(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 October 2001 (18.10.2001)

PCT

(10) International Publication Number
WO 01/77295 A1

(51) International Patent Classification[7]: C12N 1/16, 15/80, A61K 48/00

(21) International Application Number: PCT/GB01/01626

(22) International Filing Date: 11 April 2001 (11.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0008748.6 11 April 2000 (11.04.2000) GB

(71) Applicant *(for all designated States except US)*: **F2G LIMITED** [GB/GB]; The Manchester Incubator Building, Grafton Street, Manchester M13 9XX (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **DENNING, David, Wemys** [GB/GB]; 40 Crescent Road, Hale, Cheshire WA15 9NA (GB). **BROOKMAN, Jayne, Louise** [GB/GB]; 7 Fernhill Close, Glossop, Derbyshire SK13 7BE (GB). **RICKERS, Andre** [GB/GB]; 9 Moorfield Street, Hollingworth, Cheshire SK14 8JU (GB). **BIRCH, Mike** [GB/GB]; 2 Crombouke Fold, Ellenbrook, Worsley, Manchester M28 1ZE (GB).

(74) Agents: **WALSH, David, Patrick** et al.; Appleyard Lees, 15 Clare Road, Halifax, West Yorkshire HX1 2HY (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/77295 A1

(54) Title: MUTANT BANK

(57) **Abstract:** The invention relates to a mutant bank of diploid micro-organisms which consists of a population of mutant cells in which at least one cell has a random mutation which disrupts the activity of at least one gene, wherein the micro-organism is inducible into haploid form. The invention further relates to a method of using the mutant bank to identify the genes which contribute to a chosen phenotype.

MUTANT BANK

The present invention relates to the generation of mutant banks of micro-organisms comprising a population of mutant cells in which essentially every gene within the genome is inactivated and uses thereof.

Due to the advances made by various genome sequencing projects, the art has reached the position that many genes have now been cloned and sequenced. However, a problem remains that for most identified genes a link has not been made with function (i.e. how an identified gene may effect phenotype).

One way by which gene function may be identified is to compare the phenotypes of mutant cells or organisms with the known phenotype of a "wild-type" cell or organism. The gene responsible for this phenotypic change may be identified by DNA sequencing of mutant and wild-type cells. The identity of the gene product may then be easily established. Such knowledge can give rise to industrially useful processes in which the activity of the gene, or its product, is modulated to achieve specific goals. For instance, identification of a mutation that reduces the growth of a pathogenic mould provides valuable information for developing therapeutic strategies.

Various strategies have been proposed for identifying mutations in specific organisms. For instance, an insertional mutation technique for plants is disclosed in WO 99/14373. This document discloses a means of selecting and identifying insertional mutations in a population of plant cells based upon a method utilising the Ti plasmid of *Agrobacterium tumefaciens* (a gram-negative soil bacterium) for effecting transformation.

EP-A-0 870 835 also relates to the use of the Ti plasmid of *Agrobacterium*. It discloses that the abovementioned Ti plasmid technique may also be used to transform moulds from the fungal subdivisions Ascomycotina, Basidiomycotina, Deuteromycotina, Mastigomycotina and Zygomycotina.

One of the problems associated with the mutant populations disclosed by the prior art is that often no account is taken of the effect of lethal mutations. A lethal mutation results in the death of particular individuals and therefore any mutant population generated according to such prior art techniques would result in an "incomplete" population comprising "non-lethal" mutants. It is an aim of the present invention to obviate or mitigate this disadvantage, as well as others, associated with the prior art.

According to a first aspect of the present invention there is provided a mutant bank of micro-organisms comprising a population of mutant cells in which at least one cell has a mutation that disrupts the activity of at least one gene, said micro-organisms being diploid and being inducible into haploid form.

The micro-organisms may normally occur in a haploid form and are, preferably, first induced into the diploid form.

Preferably, a plurality of cells in the population each individually have a mutation which disrupts the activity of at least one gene, so that, preferably, collectively the said plurality of cells have mutations in a plurality of genes within the genome.

Preferably, the said plurality of genes makes up 0.001% of the genes within the genome, more preferably 0.01%, most preferably 0.1%, even more preferably, 10%, even more especially substantially all the essential genes and most especially all the genes of the genome of the micro-organism.

According to a second aspect of the present invention there is provided a method of generating a mutant bank of micro-organisms comprising a population of mutant cells in which at least one cell has a mutation that disrupts the activity of at least one gene, said method comprising the steps of:-

(i) culturing a population of micro-organisms, and

(ii) inducing a mutation in at least one cell of the population which mutation disrupts the activity of at least one gene in the genome of the micro-organism, said micro-organisms being diploid and being inducible into haploid form.

Preferably, the method comprises the further step of exposing the diploid micro-organisms to an agent that induces the micro-organisms into haploid form after generating the mutant bank.

Preferably, the method comprises the further step of separating and culturing the haploid micro-organisms as single clones after the induction of the micro-organisms into haploid form.

Preferably, the method comprises the further step of selecting clones for which the chosen phenotype is altered relative to a wild type micro-organism after separating and culturing the haploid micro-organisms.

Preferably, the method comprises the further step of identifying the mutated gene in each of the selected clones after selecting clones for which the chosen phenotype is altered relative to a wild type micro-organism.

The micro-organisms may normally occur in a haploid form. Therefore, preferably, the method of the second aspect further includes the step of first inducing the haploid into diploid form prior to culturing the said population.

According to a third aspect of the present invention there is provided a method of identifying genes in a micro-organism which contribute to a chosen phenotype comprising:-

(i) generating a mutant bank of diploid micro-organisms consisting of a population of mutant cells in which at least one cell has a mutation which disrupts the activity of at least one gene;

(ii) exposing the diploid micro-organisms to an agent that induces the micro-organisms into haploid form;

(iii) separating and culturing the haploid micro-organisms as single clones; and

(iv) selecting clones for which the chosen phenotype is altered relative to a wild type micro-organism; and

(v) identifying the mutated gene in each of the selected clones.

According to a fourth aspect of the present invention there is provided a mutant bank of diploid micro-organisms consisting of a population of mutant cells in which each individual cell has a mutation that disrupts the activity of one gene, said population collectively having a mutation in every gene within the genome and wherein the mutant bank may be induced into haploid form.

According to a fifth aspect of the present invention there is provided a method of identifying genes in a micro-organism which contribute to a chosen phenotype comprising:

(i) generating a mutant bank of diploid micro-organisms consisting of a population of mutant cells in which each individual cell has a mutation which disrupts the activity of one gene, said population collectively having a mutation in every gene within the genome;

(ii) exposing the diploid micro-organisms to an agent that induces the micro-organisms into haploid form;

(iii) separating and culturing the haploid micro-organisms as single clones; and

(iv) selecting clones for which the chosen phenotype is altered relative to a wild type micro-organism; and

(v) identifying the mutated gene in each of the selected clones.

By "disrupts the activity of one gene" we mean that the gene contains a mutation that prevents transcription or translation of the protein encoded by the gene. Alternatively, the translated protein has no activity, or at least altered activity, relative to the wild type gene product that results in a measurable phenotypic change.

The mutant banks may be used according to the present invention to screen the genome of the micro-organism for genes that are linked with a specific phenotype.

This knowledge may be used to develop modulators of the gene and may lead to the development of new medicaments, pesticides, disinfectants etc. which may be targeted against the micro-organism.

The prior art does not contemplate mutant banks of micro-organisms comprising a population of mutant cells in which essentially the activity of every gene within the genome is disrupted and which may be converted between diploid and haploid forms.

One of the advantages of the mutant banks according to the invention is that a complete diploid population may be cultured under suitable conditions. This is possible because the wild type copy of the gene will rescue any individual that may otherwise grow poorly, or even not be viable, should the mutant copy of the gene predominate. Thus, the diploid mutant bank may comprise a population of mutant, viable cells in which essentially the activity of every gene within the genome is disrupted rather than an "incomplete" population comprising "non-lethal" mutants.

Furthermore, the mutant bank may be comprised within a population of cells which also comprises cells in which no mutations may have occurred and/or in which more than one mutation may have occurred.

When the diploid mutant bank is induced into haploid form (step (ii)) of the third and fifth aspects any phenotypic changes in the haploid mutants may be interpreted by a technician to provide valuable information about gene function. For instance, the haploid cells may be grown under normal culture conditions. Under these circumstances, the death of a haploid clone would imply that the mutation caused a lethal phenotypic change. The clone could then be isolated from the original diploid sample and the mutant gene identified.

When organisms are used which are naturally found as haploids, it is preferred that the method of the third and fifth aspects of the invention comprises an initial step

of converting the haploid cells into diploid form before mutations are induced as defined by step (i) of the method of the third and fifth aspects of the invention.

Organisms which are normally haploid, may be induced into diploid form utilising the methods described in the Examples at 1.4 and 1.5.

It is preferred that the mutant banks are used according to the invention to screen the genome of a micro-organism for genes which contribute to a chosen phenotype by growing the micro-organisms in permissive or selective growth media. The choice of permissive or selective growth media will depend upon the phenotype of interest. For instance, the haploid cells may be initially grown in a media which is osmotically buffered such that cells with weak cell walls (which would lyse in a normal media) are able to survive. Clones which survive under such circumstances, and which do not when grown in normal media, may be selected. The mutant genes, (and thereby the gene products), in the selected clones may be identified by conventional molecular biology techniques. The identification of such gene products may be used as an aid for rationally identifying gene products which regulate osmoregulation and could also identify leads for developing anti-microbial agents that will induce cell lysis by disrupting osmoregulation.

Mutant banks used according to the invention are particularly useful for identifying new anti-microbial drug targets. By a "drug target" we mean the site of action of a drug.

By identifying essential genes in the mutant banks and then analysing them (e.g. using computer analysis - bioinformatics), the commonality of the gene in other micro-organisms (or fungi) can be established and any significant differences between mammals (e.g. man or mouse) and the micro-organism defined. For instance, a good antifungal drug target would be a protein encoded by a gene which is shared by all fungi (and be very similar between different species) and not found in man at all. By a continuous process of identifying possible new targets, on the basis that an identified

mutant gene had an effect on phenotype, and then bioinformatic analysis to rule candidates out, multiple new targets may be identified in a short time-frame.

The mutant banks also provide a simple means of establishing the mechanism of action of a compound known to have anti-fungal activity. Colonies from the banks may be plated out on media containing the anti-fungal compound of interest and resistant or less susceptible colonies examined to determine which gene has been inactivated by insertional mutagenesis. This gene product is the likely target for that anti-fungal agent. This represents a rapid and simple means for determining the mechanism-of-action of the agent.

In addition to identifying and validating anti-fungal targets, other special characteristics can be selected for. For instance, the transporter genes that mediate drug resistance may be identified according to the method of the invention. There are about 30 transporter genes in yeast but only 3 have been found in *Aspergillus fumigatus* to date. In fungi the most common resistance mechanism is due to the active export of potentially lethal drugs from within the fungal cell. Transport mutants may be identified by selecting from the haploid mutant bank for drug sensitive strains (indicating that a transporter gene is non-funtional). Drug sensitive mutants (i.e. incapable of drug export) will make good strains for drug development because the potential for future resistance by these efflux mechanisms can be directly assessed.

The micro-organism may be a yeast or any fungus, preferably, a filamentous fungus which may be switched between diploid and haploid form.

A most preferred micro-organism is *A. fumigatus* and related strains. *A. fumigatus* is a haploid organism and the most common human mould pathogen. The genome of *A. fumigatus* is thought to comprise 8,000-11,000 genes. It is important when developing anti-microbial agents to select a suitable target which will not only kill the micro-organism but which are also selective. In this respect it is expected that approximately 750-1000 genes in *A. fumigatus* will be essential and therefore potential targets. Of the essential genes possibly a quarter will be new genes with no known

counterpart in yeast-fungi, over 100 are likely to be structural genes (which do not make good targets) and as many as 500 will be common to man. This leaves an estimated 150 candidate targets which may encode proteins that could represent leads for developing anti-microbial agents. The method of the second aspect of the invention may be used to identify candidate clones. Further screening of the candidates may be achieved by employing bioinformatics and it is estimated that such an approach will generate 15-25 validated targets. Some may be shared by bacteria, raising the possibility of finding anti-microbials that have both antibacterial and antifungal activity.

A. fumigatus clinical isolates AF300 and AF293 (available to the public from the NCPF repository (Bristol, U.K.) and the CBS repository (Belgium)) are preferred strains which may be formed into mutant banks according to the present invention.

Another preferred micro-organism is *Candida glabrata*. *Candida* species are important yeast fungal pathogens of humans. Until recently, *C. glabrata* was considered a relatively non-pathogenic commensal fungal organism of human mucosal tissues. However, with the increased use of immunosuppressive agents, mucosal and systemic infections caused by *C. glabrata* have increased significantly, especially in the human population infected with HIV. *C. glabrata* currently ranks second or third as the causative agent of superficial (oral, oesophageal, vaginal, or urinary) or systemic candidal infections. A number of factors have been proposed as being important to virulence but details of the host-pathogen interaction are, however, largely unknown. A major obstacle in *C. glabrata* infections is their innate resistance to azole antimycotic therapy, which is very effective in treating infections caused by other *Candida* species. *C. glabrata*, formerly known as *Torulopsis glabrata*, contrasts with other *Candida* species in its nondimorphic blastoconidial morphology and haploid genome. We have realised that the possession of a haploid genome is a useful feature in the production of mutant banks as most *Candida* species are obligate diploids with no known haploid forms and therefore *C. glabrata* is a preferred micro-organism which may be made into a mutant bank according to the invention.

A. fumigatus and *C. glabrata* normally have only one set of chromosomes (haploid) and conventionally would not be considered for the formation of a bank of mutants because lethal gene knockouts would not survive and therefore the bank would be incomplete. However, the inventors have appreciated that these moulds can be induced to carry a double set (diploid) of genes. They therefore realised that this property could be exploited such that a stock mutant bank may be maintained in diploid form. One gene from each pair should be in wild type form and thereby allow the organisms to be phenotypically normal. When desired the cells may be allowed to return to haploid form and the effects of the mutant gene identified.

It will be appreciated that the method of the invention used in *A. fumigatus* may be usefully employed to gain knowledge relating to homologous genes for other species of micro-organism. For instance, biotechnology companies are interested in improving strains of filamentous fungi which are used extensively in industry for the production of soy sauce, citric acid etc. Furthermore, the production of mutant banks or knowledge gained from the *Aspergillus* mutant bank in plant pathogenic fungi is useful to the agrochemical industry for the development of broad-specificity anti-fungal drugs.

Preferably, the mutant banks comprise a population of cells in which the mutations have been randomly induced. Preferably, there has been a random knock out of essentially all genes in the micro-organism.

The mutant banks may be generated (according to step (i) of the third and fifth aspects of the invention) by exposing a population of micro-organisms to an agent which randomly causes single mutations in each individual micro-organism. It is preferred that the mutant bank is formed by an insertional mutagenesis method. Preferably, a DNA molecule is inserted into each gene to cause the mutation. The DNA molecule may be a selectable marker which is, preferably, a *pyrG* gene, more preferably, the *pyrG* gene from *A. fumigatus*.

Advantageously, the *pyrG* gene is a preferred selectable marker because it may be very tightly regulated since there is an absolute requirement of *pyrG* mutants for uridine and uracil supplementation in the growth medium. Advantageously, this substantially reduces the number of false negatives.

Preferably, the *pyrG* gene is harboured on a plasmid which is, preferably, an *Aspergillus pyrG* containing plasmid.

The micro-organism may be a fungus, preferably, a filamentous fungus and, most preferably, *A.fumigatus*. In a preferred embodiment the micro-organism is AF300 or AF293.

Alternatively, the micro-organism may be a yeast, preferably, *C.glabrata*.

Preferably, the mutations have been randomly induced, more preferably, by an insertional mutagenesis method. Preferably, a DNA molecule is inserted into each gene to cause the mutation. The DNA molecule may be a selectable marker which is, preferably, a *pyrG* gene, more preferably, the *pyrG* gene from *A. fumigatus*. Preferably, the *pyrG* gene is harboured on a plasmid which is, preferably, an *Aspergillus pyrG* containing plasmid.

Examples of the *pyrG* containing plasmid are shown in Figures 6, 7, 10 and 11. Preferably, the plasmid is introduced into the micro-organism by electroporation (see section 1.7.1).

Preferably, the diploid micro-organism is a mutant for the *pyrG* phenotype.

Advantageously, this allows for the selection of the *pyrG* gene containing plasmids.

Preferably, the diploid mutant micro-organism is generated using the following steps:-

(i) isolating a first haploid mutant comprising a first auxotrophic marker wherein the first auxotrophic marker results in the *pyrG*⁻ phenotype;

(ii) isolating a second haploid mutant using the first haploid mutant isolated in step (i), said second mutant comprising the first *pyrG*⁻ auxotrophic marker and a second auxotrophic marker;

(iii) isolating a third haploid mutant using the first haploid mutant isolated in step (i), said third mutant comprising the first *pyrG*⁻ auxotrophic marker and a third auxotrophic marker; and

(iv) mating the second and third haploid mutants isolated in steps (ii) and (iii) to generate the diploid mutant exhibiting the *pyrG*⁻ phenotype.

Preferably, the second and third auxotrophic markers are selected from *niaD*⁻ or *cnx*⁻. Preferably, cells which are unable to utilise nitrate alone were classified as *niaD*⁻ mutants and, preferably, mutants unable to utilise mitrate or hypoxanthine were classified as *cnx*⁻ mutants.

Preferably, the mutant diploid micro-organism is a double mutant for the *pyrG*⁻ phenotype, i.e. *pyrG*⁻/*pyrG*⁻ phenotype. Preferably, the double mutants (*pyrG*⁻/*pyrG*⁻ phenotype) were made using a haploid single *pyrG*⁻ mutant as a parental strain. Preferably, the order of making the diploid double mutants was found to be important since starting with a *niaD*⁻ or *cnx*⁻ mutant as the parental strain we were unable to then isolate a *niaD*⁻/*pyrG*⁻ or *cnx*⁻/*pyrG*⁻ mutant.

Preferably, the method comprises use of the Frontier method (section: 1.4.1) for the production of diploid micro-organisms in *A.fumigatus*.

Preferably, the haploid mutants are incubated at a temperature in the range of 20-36°C, more preferably, in the range of 24-33°C and, most preferably, in the range of 26-30°C. Most preferably, the haploid mutants are incubated at 28°C.

Advantageously, incubation at a lower temperature slows down growth of the haploid mutants thereby increasing the chance of heterokaryon formation and the subsequent production of diploid mutants with which to make the mutant bank.

Preferably, putative mutant cell transformants are identified using a microscope when they are substantially not visible to the naked eye at the earliest stage possible in their development. Preferably, putative mutant cell transformants are identified at 32 to 48 hours in to their development, more preferably, at 34 to 46 hours and, most preferably, at 36 to 44 hours in to their development.

Preferably, the minimum feasible time that putative mutant cell transformants could be identified using a microscope is 36-48 hrs. Most methods rely on the macroscopic identification of colonies, i.e. by eye when they have reached a particular size. By this time, many of the colonies are sporulating which poses a cross-contamination problem with other nearby putative transformants. More importantly, due to the randomness of the transformation procedure, there will be a whole range of transformants which are affected to a greater or lesser degree in their growth rates. Some of these transformants may never reach macroscopic size and would hence be lost from the population of the mutant bank as a whole. Identification of these transformants by microscope at an early stage allows them to be transferred to a richer, more complete medium where they have a better chance to grow to a size where DNA can be extracted and the genetic mutation identified.

According to a still further aspect of the present invention there is provided SEQ ID No. 27 and homologues thereof, preferably, functional and/or structural homologues thereof.

SEQ ID No. 27 (see sequence listing) may be used for the manufacture of a medicament for treatment of an infection of *A.fumigatus*.

In an alternative emodiment, the insertional mutagenesis may be carried out using the Ti plasmid. The use of the Ti plasmid in filamentous fungi has previously been described (de Groot *et al.* (1998) Nature Biotechnology 16:839-842), but it has not been used for the generation of mutant banks according to the first aspect of the invention nor has it been used in filamentous fungi such as *A. fumigatus* or in *C. glabrata.*

The Ti plasmid transformation methods disclosed in EP-A-0 870 835 (incorporated herein by reference) may be used, and adapted as appropriate, to generate mutant banks according to the invention. According to one embodiment (see the example), the physical steps of transforming diploid *A. fumigatus* may be essentially the same as those disclosed in EP-A-0 870 835.

The mutant bank may be generated by transformation with a Ti plasmid vector based on the binary vector pBIN19 (disclosed in: Bevan. (1984) Nucleic Acids Research 22: 8711-8721).

It is most preferred that LBA4404 and GV3101 strains of *A. tumefaciens* (identified in Table 1) are used to generate the mutant banks of the present invention.

TABLE 1

Strain	Genotype	Reference disclosing strain
LBA 4404	Ach5 Rif ^{r} containing plasmid pAL4404 (ΔT_L, ΔT_R, Δtra, Δocc) a deletion of pTiAch5.	Oooms *et al.* (1981) Gene **14**: 33-50.
GV3101	C58C1 Rif ^{r} containing plasmid pMP90 (Gm ^{r}) a deletion of pTiC58.	Koncz & Schell, (1986) Molecular and General Genetics **204**: 383-396.

Most preferred transformation steps are described in 1.7 of the Example.

Following generation of the diploid mutant bank, the cells are induced into haploid form according to step (ii) of the method of the invention. A preferred technique for inducing the cells into haploid form is described in the Example under 1.6.

The haploid cells from step (ii) are then separated and cultured according to step (iii). This may be achieved by conventional dilution and spread plating of the culture and is also described at 1.6 of the Example.

Clones with an altered phenotype (step (iv)) may be isolated by a variety of conventional means (e.g. by growth on a selective medium).

Once clones have been selected according to step (iv) of the method of the invention it is preferred that the mutated gene is identified. Identification of the mutated gene may be achieved by including sections of marker DNA during the generation of the mutant bank. For instance, a gene may be insertionally inactivated by incorporating a marker DNA sequence which may be later identified using conventional molecular biology techniques (e.g. using labelled probes for the marker). Alternatively, the marker may be used as a target for a primer which can be used to directly sequence and amplify the mutant gene. This approach has several advantages over known mutational techniques. For instance:

(1) it is not necessary to have discovered the gene previously;

(2) the essential function of the gene is established concurrently with its identification; and

(3) it is extremely rapid.

Various methods known to those skilled in the art may be used to identify the gene. Preferred methods are outlined at 1.9 and 1.10 of the methods section of the Example.

All of the features described herein may be combined with any of the above aspects, in any combination.

An embodiment of the present invention will now be described, by way of example, with reference to the accompanying drawings in which:-

Figure 1 is a schematic representation of plasmid pAN7-1;

Figure 2 is a schematic representation of plasmid pRok2;

Figure 3 is a schematic representation of plasmid pRic1;

Figure 4 is a schematic representation of plasmid pUC18;

Figure 5 is a schematic representation of plasmid pRic2;

Figure 6 is a schematic representation of plasmid pRic3;

Figure 7 is a schematic representation of plasmid ppyrG;

Figure 8 is a schematic representation of plasmid pCR2.1;

Figure 9 is a schematic representation of plasmid pTT;

Figure 10 is a schematic representation of plasmid pMB2;

Figure 11 is a schematic representation of plasmid pMB3;

Figure 12 is a BLASTX search result used for the identification of mutated gene in strain MA205;

Figure 13 is a Southern blot showing integration events; and

To the Sequence Listing.

EXAMPLE

1. METHODS

1.1 Bacterial and Fungal Strains:

For bacterial cloning the *E. coli* strain Top10 (F$^-$, *mcr*A Δ (*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔM15 Δ*lac*X74 *deo*R *rec*A1 *ara*D139 Δ(ara-leu)7697 *gal*U *gal*K *rps*L(Str^R) *end*A1 *nup*G) was used.

The *A. tumefaciens* strains LBA4404 and GV3101 (see Table 1) were used for the transformation of *A. fumigatus*.

A. fumigatus clinical isolates AF300 and AF293 (available to the public from the NCPF repository (Bristol, U.K.); the CBS repository (Belgium) or from Dr. David

Denning clinical isolate culture collection, Hope Hospital, Salford. U.K.) are preferred strains which may be formed into mutant banks according to the present invention.

AF300: Isolated 1995. Royal Manchester Children's Hospital. Leukaemia patient.

AF293: Isolated 1993. Donated by Shrewsbury PHLS. Lung biopsy, invasive aspergillosis with aplastic anaemia.

1.2 Plasmid construction:

1.2.1 Plasmid pRic1

Plasmid pRic1 (see Figure 3) was constructed by cloning a 4 kb *Hind*III – *Bgl*II fragment, which is present on the vector pAN7-1 (Punt *et al.*(1987) Gene 56: 117-124; see Figure 1) and contains the promoter from the *Aspergillus nidulans gpd* gene fused to the coding region of the *E. coli* hygromycin B phosphotransferase (*hph*) gene and followed by terminator sequences from the *A. nidulans trpC* gene, into the *Hin*dIII – *Bam*HI site of the Cauliflower Mosaic Virus (CaMV) 35s promoter region of the binary vector pRok2 (a derivative of the binary vector pBin19) pRok2 is disclosed in: Baulcombe *et al.* (1986) Nature **321**: 446-449; and Figure 2).

1.2.2 Plasmid pRic2

Plasmid pRic2 (see Figure 5) was constructed by cloning the entire pUC18 plasmid (see: Yanisch-Perron *et al.* (1985) Gene **33**: 103-119; and Figure 4) linearised with *Hin*dIII into the *Hin*dIII site of pRic1.

1.2.3 Plasmid pRic3

Plasmid pRic3 (see Figure 6) was constructed by cloning the entire ppyrG plasmid (Fungal Genetics Stock Centre, http://www.fgsc.net/; and Figure 7) linearised with *Xba*I into the *Xba*I site of pRok2.

The *A. tumefaciens* strain GV3101 was electroporated with the constructs pRic1, pRic2 and pRic3 using a Biorad Genepulser (2.5 kV, 600 Ω, 25 μF). The constructs

pRic1, pRic2 and pRic3 were transformed into *A. tumefaciens* strain LBA4404 using a triparental mating method as follows. Freshly grown cultures of *E. coli* Top10 containing pRic1, pRic2 or pRic3, *E. coli* Top10 containing the helper plasmid pRK2013 and *A. tumefaciens* LBA4404 were mixed in the middle and streaked to the edges of a LB plate which was then incubated overnight at 25 °C. A loop of cells from the LB plate was then streaked onto a fresh LB agar plate containing rifampicin (200 μg/ml) and kanamycin (50 μg/ml) for pRic1 transformations and rifampicin (200 μg/ml), kanamycin (50 μg/ml) and ampicillin (100 μg/ml) for pRic2 and pRic3 transformations. This plate was incubated at 25 °C until single colonies started to form (these single colonies represent *A. tumefaciens* LBA4404 containing pRic1, pRic2 or pRic3).

1.2.4 Plasmids pMB2 and pMB3

The homologous *pyrG* gene from wild type AF293 was amplified using primers designed to the start of the promoter and end of the terminator of the *A. fumigatus pyrG* gene (Weidener, G. and d'Enfert, C. (1998) Current Genetics 33, 378-385).

Primers AFpyrG5 (5'-cta cct cga gaa ttc gcc tca aac-3'; SEQ ID No 25) and AFpyrG3 (5'-ggc gac gaa ttc tgt ctg aga g-3'; SEQ ID No 26) were used in PCR reactions containing *Taq* polymerase (ABGene) and Expand *Pfu* polymerase (Roche Molecular Systems).

Reactions which amplified the predicted 1.9 kb fragment were cloned directly. Where *Taq* polymerase was used, fragments were cloned directly into PCR cloning vector pTT (Genpak Ltd, see Figure 8). Where Expand *Pfu* was used, fragments were first treated with *Taq* polymerase for 10 min at 72 °C. Reactions were then passed through a cleanup column (Qaigen Ltd) to remove residual enzymes and nucleotides and then cloned into vector pCR2.1 (Invitrogen Ltd, see Figure 9) according to the manufacturers instructions.

In both cases ligated vector/insert were transformed into Top10 electrocompetent cells (Invitrogen Ltd) and plated onto LB agar containing ampicillin (100 μg ml). Following overnight incubation at 37 °C six individual colonies from each reaction were sub-cultured into LB broth containing ampicillin (100 μg ml). After overnight incubation at 37 °C, plasmids were extracted using Qaigen Spin Mini plasmid extraction kits and digested with *Eco*RI.

Plasmids containing the predicted 1.9 kb insert released by *Eco*RI digestion were deemed positive clones. Aliquots of selected positive clones were sequenced to confirm the presence of the *A. fumigatus pyrG* gene. Sequencing was performed by MWG Biotech UK Ltd (Waterside House, Peartree Bridge, Milton Keynes, MK6 3BY).

Plasmid pMB2 (see Figure 10) was generated from a *Taq* polymerase clone inserted into pTT. Plasmid pMB3 (see Figure 11) was generated from a *Pfu* polymerase clone inserted into pCR2.1.

1.3 Isolation of *A. fumigatus* haploid auxotrophs

1.3.1 Nitrate assimilation mutants (*niaD*⁻ and *cnx*⁻)

Haploid strains of *A. fumigatus* AF293 were identified which were deficient in the utilisation of nitrogen in the form of nitrate. Nitrate mutants were selected after inoculation of AF293 conidia (10^7 per plate) on chlorate plates (600mM sodium chlorate) supplemented with 10mM asparagine. Colonies growing on chlorate plates were characterised further by their ability to grow on agar containing either 10mM nitrate, 10 mM nitrite or 10mM hypoxanthine as the sole nitrogen source. Colonies unable to utilise nitrate alone were classified as *niaD*⁻ mutants. Colonies unable to utilise nitrate or hypoxanthine were classified as *cnx*⁻ mutants. Reversion rates were checked by plating conidia onto nitrate media at 10^7 conidia/ml. No reversion was seen at this level.

1.3.2 *pyrG*⁻ mutants

The *pyrG* phenotype is due to a mutation in the gene encoding orotidine 5-phosphate decarboxylase. This causes a requirement for supplementation of the growth media with 10mM uracil and 10 mM uridine in order to allow *pyrG* mutants to grow.

Generation of *pyrG* mutants is achieved by the use of the metabolic inhibitor fluoroorotic acid (FOA). FOA is added to complete media (CM agar) at a concentration of 1 mg/ml. CM contains in grams per litre: malt extract, 20; glucose, 10; peptone, 1; agar, 15.

A. fumigatus spores are spread onto CM/FOA plates at a concentration of 10^7 and 10^8 per plate. Plates are incubated at 37 °C until colonies appear (usually 3-5 days). Colonies are then picked onto CM/FOA replica plates for purity and to prepare for confirmation of the nutritional status.

Confirmation of a *pyrG* phenotype is achieved by checking the requirement for uracil/uridine in the growth media. Two sets of minimal media (MM) are prepared, one supplemented with uracil/uridine (10mM each final concentration) and one with no supplementation. MM contains in grams per litre: $NaNO_3$, 6.0; KH_2PO_4, 1.52; KCl, 0.52; $MgSO_4.7H_2O$, 0.52; glucose, 10.0; trace element solution, 1.0 ml; media is adjusted to pH 6.5 with KOH. Trace element solution contains in grams per litre: $FeSO_4$, 1.0; $ZnSO_4.7H_2O$, 8.8; $CuSO_4.5H_2O$, 0.4; $MnSO_4.4H_2O$, 0.15; $Na_2B_4O_7.10H_2O$, 0.1; $(NH_4)_6Mo_7O_{24}.4H_2O$, 0.05. Uracil is quite insoluble and is made at 20mM stock in water, uridine is made up at 500mM stock in water. Both solutions are filter sterilised. CM and MM are made up and autoclaved in small volumes in order to compensate for the large volume of uracil which needs to be added (due to its low solubility).

Suspected *pyrG* mutants are replica plated onto the two agars and incubated at 37 °C. True *pyrG* mutants should emerge within 48 h on plates supplemented with uracil/uridine and no growth should occur on the corresponding plate containing no supplementation.

1.3.3 Isolation of haploid double auxotrophs (*pyrG⁻/niaD⁻* and *pyrG⁻/cnx⁻*)

The isolation of haploid double auxotrophs (*pyrG⁻/niaD⁻* and *pyrG⁻/cnx⁻*) is necessary for the production of diploids containing a double *pyrG⁻/pyrG⁻* phenotype. The resulting *pyrG⁻/pyrG⁻* diploids can then be used in transformation experiments utilising plasmids pRic3, ppyrG, pMB2 and pMB3 all of which use the *pyrG* gene as a marker for DNA integration.

Using a haploid *pyrG⁻* mutant (produced as in section 1.3.2) as the parental strain the exact protocol was followed as in section 1.3.1 (Nitrate assimilation mutants (*niaD⁻* and *cnx⁻*)). In this case all media used for the selection and confirmation of nitrate assimilation mutants was supplemented with uracil/uridine (10mM each final concentration) due to the presence of the *pyrG⁻* phenotype.

The order of making the double auxotrophs was found to be important. Starting with a *niaD⁻* or *cnx⁻* mutant as the parental strain we were unable to then isolate a *niaD⁻/pyrG⁻* or *cnx⁻/pyrG⁻* double mutant. All our double mutants were made using a haploid *pyrG⁻* mutant as the parental strain.

1.4 Production of diploids in *A. fumigatus*:

These methods are based around a paper published by Strømnæs and Garber in 1963 (Strømnæs, Ø and Garber, E.D. (1963) Genetics **48**, 653-662). Using the following methods diploid strains can be produced which are either (*niaD⁻/cnx⁻*) when using single auxotrophic markers or (*pyrG⁻/pyrG⁻/niaD⁻/cnx⁻*) when using double auxotrophic markers. Diploids with the *niaD⁻/cnx⁻* phenotype are used in transformation experiments which utilise hygromycin resistance as a marker for DNA integration. Diploids with the *pyrG⁻/pyrG⁻/niaD⁻/cnx⁻* phenotype are used in transformation experiments which utilise the *pyrG* gene as a marker for DNA integration.

1.4.1 Frontier method

This is our method of choice for the production of diploids in *A. fumigatus*. Spores from *niaD⁻* and *cnx⁻* mutants, or *niaD⁻/pyrG⁻* and *cnx⁻/pyrG⁻* mutants (both at

10^6 conidia per ml) were each inoculated on to one half of an agar plate containing Vogel's agar (Vogel, H.J. (1956) which is a convenient growth medium for *Neurospora* (medium N) Microbiol. Gen. Bull. 13: 42-44). Plates were incubated at 28 °C until the colonies had developed sufficiently to merge. Conidia were collected from mycelia at the junction of the two colonies. Conidia were plated onto media containing nitrate (10mM) as the sole nitrogen source and incubated at 37 °C until colonies had developed. Colonies growing at this stage were regarded as presumptive diploids. The leading edge of growing colonies were transferred to fresh nitrate media and allowed to develop. Conidia from these colonies were then transferred to nitrate media and incubated at 37 °C.

During the initial step of this method the plates are incubated at 28 °C as opposed to 37 °C. Incubation at the lower temperature slows down the growth rate, increasing the chance of heterokaryon formation and the subsequent production of diploids.

When double auxotrophs were used to produce diploids all media was supplemented with uracil/uridine (10mM each final conc) due to the presence of the *pyrG*⁻ phenotype.

1.4.2: Mat method

Conidia from both *niaD*⁻ and *cnx*⁻ mutants, or *niaD*⁻/*pyrG*⁻ and *cnx*⁻/*pyrG*⁻ mutants were collected and counted. Fifty µl of conidia (approximately 1 x 10^6 per ml) from each mutant phenotype was mixed with 200 µl of complete medium (CM broth) and incubated overnight at 37 °C. CM broth contains in grams per litre of water: malt extract, 20; bacto-peptone, 1; D-glucose, 20 (Rowlands & Turner. (1973) Molecular and General Genetics **126**: 201-216.). The resulting mycelial mat was placed on minimal media containing 10mM nitrate as the sole nitrogen source and incubated at 37 °C until growth had occurred. Subcultures of any outgrowths of the primary inoculum were taken by removing sections of the leading edge and transferring to fresh nitrate medium. Growth on this media indicates a stable diploid colony.

The ability to grow on nitrate media indicated the formation of a diploid colony since neither the *niaD*⁻ or *cnx*⁻ (nor *niaD*⁻/*pyrG*⁻ or *cnx*⁻/*pyrG*⁻) mutants can grow on this media. Therefore, any colony growing on nitrate must have a functional gene i.e. obtained through genetic fusion of the two mutants.

When double auxotrophs were used to produce diploids all media was supplemented with uracil/uridine (10mM each final conc) due to the presence of the *pyrG*⁻ phenotype.

1.4.3: Protoplast Fusion

Conidia (10^9 total) from *niaD*⁻ and *cnx*⁻ mutants, or *niaD*⁻/*pyrG*⁻ and *cnx*⁻/*pyrG*⁻ mutants were inoculated into 100 ml Vogel's medium and incubated for 16 h at 37 °C. Germlings were harvested by filtration and suspended in 50 ml protoplasting buffer, PB (1M NaCl, 10mM CaCl) containing lysing enzyme at 2mg/ml. Cultures were incubated at 30 °C with constant shaking (100 rpm) for 30-90 min. Protoplast formation was monitored microscopically until around 80-90% of germlings were converted to protoplasts. Protoplasts were harvested by filtration, concentrated by centrifugation (2000g for 5 min) washed in PB and resuspended in sorbitol buffer (0.9M sorbitol, 0.125M EDTA pH 7.5). Protoplasts were diluted to a concentration of 10^7 per ml. Equal amounts (100 µl) of protoplasts from both *niaD*⁻ and *cnx*⁻ mutants, or *niaD*⁻/*pyrG*⁻ and *cnx*⁻/*pyrG*⁻ mutants were mixed with 100 µl of 30% PEG and placed on ice for 20 min.

100 µl and 50 µl aliquots were added to tempered nitrate media (10mM nitrate final) supplemented with 1.2M sorbitol and 10mM CaCl. Plates were incubated at 30 °C until colonies emerged. Colonies were transferred to fresh nitrate medium and allowed to develop.

When double auxotrophs were used to produce diploids all media was supplemented with uracil/uridine (10mM each final concentration) due to the prescience of the *pyrG*⁻ phenotype.

1.5 Production of dipl ids in *C. glabrata:*

C. glabrata is naturally haploid but diploid strains may be produced as follows.

Suitable parental *C. glabrata* haploid strains are selected which have lost the ability to make specific metabolites or enzymes. Parental strains may be selected by screening for natural spontaneous mutations or by inducing mutation through UV mutagenesis or treatment with chemical mutagens e.g. nitric acid.

Where spontaneous mutations are used, the desired mutations are screened by plating liquid cultures of *C. glabrata* onto solid complete media which may contain vitamin and mineral supplements to allow for growth of desired mutant phenotypes.

Colonies are then replica plated on to minimal media which are deficient in one particular nutrient only. Mutant auxotrophs can then be identified by comparing their growth on both complete and minimal media. Organisms which fail to grow on minimal media but grow on supplemented complete media are classified as auxtrophic for the deficient compound. It is preferable that one parent is auxotrophic for two compounds preferably from uracil, arginine, leucine, histidine and adenine and that the other parent is also auxotrophic for two of the said compounds but its auxotrophy differs from the other parent.

Formation of diploid *C. glabrata* may be achieved by joint culture of two parental auxotrophs in liquid media containing the necessary nutrients for both parents to grow. This is followed by sub-culturing organisms from this media to minimal media deficient in the auxotrophic nutrients of either parent. Diploid strains can then be isolated from this media.

Additionally diploid strains may be formed from haploid parental auxotrophs by polyethylene glycol (PEG) mediated protoplast fusion. Protoplasts from each parental haploid auxotroph are formed by digestion of the cell wall with a suitable commercial protoplasting enzyme (the preferred enzyme being Zymolase). Protoplasts from each haploid parent are mixed (the joint culture stage) in the presence of PEG.

Protoplasts are then plated onto solid minimal medium deficient in auxotrophic nutrients of the parent haploid strains but supplemented with an osmotic stabiliser to prevent protoplast bursting. Prototrophic colonies that derive from this minimal media are presumptive diploids.

1.6 Re-haploidisation

The following method was used for the re-haploidisation of diploid colonies of *C.glabrata* and *A.fumigatus*. Diploid colonies were subjected to re-haploidisation by the use of the mitotic inhibitor fluorophenylalanine (FPA).

Conidia (*A. fumigatus*) or cells (*C.glabrata*) were collected from stable diploid colonies and spread plated onto complete media containing nitrate and 0.01-0.2% FPA and incubated at 37 °C for 3 days or until rapidly growing sectors emerged (*A. fumigatus*). Conidia were collected from each sector (*A. fumigatus*) or colonies picked (*C.glabrata*) and plated onto nitrate, nitrite and hypoxanthine media and the nitrogen utilisation profiles of the resulting conidia (*A. fumigatus*) or cells (*C.glabrata*) assessed. Colonies with the nitrogen utilisation profiles of the parental strains could then be re-isolated indicating a haploid.

1.7 Transformation experiments:

1.7.1 Electroporation

This is our method of choice for the production of transformants in *A. fumigatus*.

Approximately 125 ml of YG (0.5% (w/v) yeast extract, 2% (w/v) glucose, 5mM each uridine and uracil) medium was inoculated with 10^9 conidia from an auxotrophic haploid or diploid strain. The preferred auxotrophic strains were derived from AF300 or AF293.

For transformation of swollen conidia, cultures were incubated for 4 h at 37 °C with constant shaking (200 rpm). For transformation of germlings, incubation was for 8h again at 37 °C. Swollen spores or germlings were collected by centrifugation (5 min at 5000 x g) and washed in 200 ml of ice cold sterile water. Spores or germlings were

resuspended in 10-12 ml of YED (1% (w/v) yeast extract, 1% (w/v) glucose, 20 mM HEPES, pH 8.0), and incubated at 30 °C for a further 60 min. Spores/germlings were collected by centrifugation as described and resuspended in 1 ml of EB buffer (10mM Tris pH7.5, 270 mM sucrose, 1 mM lithium acetate).

Electroporation was carried out in 50 µl aliquots of spores/germlings (5 x 10^7 germlings/conidia). 50 µl of swollen conidia/germlings were transferred to an electroporation cuvette, on ice, and 1-5 µg of transforming DNA added. This may be in the form of circular or linearised plasmid ppyrG, pMB2, pMB3 or another plasmid or DNA species carrying a complimentary DNA sequence to the auxotrophic markers of the recipient strain.

Conidia/germlings and transforming DNA were mixed and left on ice for 30 min. Conidia/germlings were electroporated in a Gene Pulser II instrument (Bio-Rad Ltd) set at 1kV, 400 Ω and 25 µF.

1 ml of cold YED was added to the cuvette and incubated at 37 °C for 1 h. Aliquots were spread on non-selective agar (Vogel's, minimal media or complete media) without urine or uracil. Colonies growing on non-selective media were deemed putative transformants.

1.7.2 PEG-mediated protoplast transformation

The following method is used for transformation of protoplasts of *C. glabrata* which are produced as described in section 1.5, and *A.fumigatus* as herein described below.

Conidia (*A.fumigatus*) from AF293 *pyrG*⁻ strains with additional *niaD*⁻ or *cnx*⁻ auxotrophy were cultured on solid CM media supplemented with 10 mM uracil and 10 mM uridine at 37 °C for 2-4 days. Conidia were collected in 10 ml PBS/0.1 % (w/v) Tween 80 and counted using a heamocytometer. 1 x 10^7 conidia were inoculated into 50 ml Vogel's media in Erlinmeyer flasks, again supplemented with 10 mM uracil and 10 mM uridine. Flasks were incubated overnight at 37 °C with constant shaking and

the resultant mycelia was harvested by vacuum filtration. Mycelia was resuspended in 20 ml protoplasting buffer (1M NaCl, 10 mM $MgCl_2$) and protoplasting enzyme (zymolase) added to a final concentration of 1-5 mg/ml. This mycelia/enzyme suspension was incubated at 30 °C with gentle shaking (80 rpm) and protoplast generation followed microscopically over a maximum 3h period. Protoplasts were harvested when most of the mycelia had been converted to protoplasts by filtration through sterile gauze.

Protoplasts were pelleted by gentle centrifugation (800 g x 5 min) and washed 2 x 10 ml in protoplasting buffer before finally being resuspended in stabilising buffer (0.9 M sorbitol, 0.1 M EDTA) to a final concentration of $2x\ 10^7$ protoplasts/ml. Up to 2 µg of linearised transforming plasmid (either ppyrG, pMB2 or pMB3) was added to the protoplast suspension and placed on ice for 30 min.

500 µl PEG 3000 (40 % (w/v)) was added dropwise to 500 µl protoplast/plasmid suspension containing 10^7 protoplasts. This PEG protoplast suspension was incubated on ice for a further 15 min before centrifugation at 100 g for 10 min. The supernatant was removed and replaced with 500 µl of stabilising buffer. Transformation reactions were either plated onto the surface of Vogel's agar plates containing 1% (w/v) glucose, 1.2 M sorbitol or mixed with 20 ml of molten tempered (to 40-45 °C) of the same agar.

Plates were incubated at 37 °C for up to 14 days and inspected for growth. Colonies growing on this selection media were deemed putative transformants.

1.7.3 *Agrobacterium*-mediated transformation

The *A. tumefaciens* strains containing the vectors pRic1, pRic2 or pRic3 were grown at 29 °C overnight on LB plates containing: rifampicin, kanamycin, ampicillin and gentamicin for *A. tumefaciens* GV3101 containing pRic2 or pRic3; rifampicin, kanamycin and gentamicin for *A. tumefaciens* GV3101 containing pRic1; rifampicin and kanamycin for *A. tumefaciens* LBA4404 containing pRic1; rifampicin, kanamycin and ampicillin for *A. tumefaciens* LBA4404 containing pRic2 or pRic3. All antibiotic

concentrations are as stated in the Plasmid construction section (1.2.3) except for gentamicin at 20 μg/ml. A single colony was streaked on a minimal medium plate. Minimal medium (MM) contains in grams per litre: K_2HPO_4, 2.05; KH_2PO_4, 1.45; NaCl, 0.15; $MgSO_4.7H_2O$, 0.50; $CaCl.6H_2O$, 0.1; $FeSO_4.7H_2O$, 0.0025; $(NH_4)_2SO_4$, 0.5; glucose, 2.0. The plates were incubated at 29 °C for 1 to 2 days.

Several colonies were inoculated in minimal medium containing the appropriate antibiotics and grown at 29 °C overnight. After dilution of *A. tumefaciens* cells to an OD_{660nm} of approx. 0.15 in induction medium the culture was grown for 6-7 hours at 29 °C. The induction medium (IM) differs from MM in that the 2 grams per litre glucose was replaced by 10mM glucose and 40mM MES (pH 5.3), 0.5% glycerol (w/v) and 200 μM acetosyringone (AS) were added. In order to confirm that the transformation of *A. fumigatus* by *A. tumefaciens* is dependent on T-DNA transfer, a negative control was included in which the *vir* inducer AS was omitted.

Conidia were obtained by growing the *A. fumigatus* strains (haploid or diploid) at 37 °C on Vogel's minimal medium agar plates for several days and subsequently washing the surface of the plates with physiological salt solution and then filtering the conidial suspension through glass wool.

For transformation of conidia, conidia were diluted in physiological salt solution at a concentration of 10^6 or 10^7 conidia per ml and 100 μl was mixed with 100 μl of the *A. tumefaciens* culture (induced as detailed using IM). Subsequently, the mixtures were plated on nitrocellulose filters placed on absorbent pads containing IM (reduced glucose concentration to 5mM) and incubated at 25 °C for 2 to 3 days. The negative control samples were incubated on IM pads in which the *vir* inducer AS was omitted. After this incubation period, the filters were transferred to Vogel's medium agar plates containing cefotaxime (200 μM) to kill the *A. tumefaciens* cells and hygromycin (400 μg/ml) to select for fungal transformants.

1.7.4 Early identification of transformants

Using the previously described transformation methods it was found beneficial to be able to identify transformants at the earliest possible stage of their development. Putative transformants were identified using a stereo microscope (Zeiss Ltd) when they were not yet visible to the naked eye and were picked using a sterile, fine gauge needle and transferred to individual petri dishes containing selective agar.

1.8 DNA isolation, PCR and Southern analysis:

To obtain mycelial material for genomic DNA isolation, approximately 10^7 *A. fumigatus* conidia were inoculated in 50 ml of Vogel's minimal medium and incubated with shaking at 200 rpm until late exponential phase (18-24 h) at 37 °C. The mycelium was dried down onto Whatmann 54 paper using a Buckner funnel and a side-arm flask attached to a vacuum pump and washed with 0.6 M $MgSO_4$. At this point it is possible to freeze-dry the mycelium for extraction at a later date. The mycelium (fresh or freeze dried) was ground to a powder using liquid nitrogen in a –20 °C cooled mortar. The powder was added to a 1.5 ml microcentrifuge tube using an ethanol-cleaned spatula (no more than 0.4 ml), 0.6 ml of extraction buffer (0.7 M NaCl; 0.1 M Na_2SO_3; 0.1 M Tris-HCl pH 7.5; 0.05 M EDTA; 1%(w/v) SDS) heated to 65° C was added and the microfuge tube was incubated at 65° C for 20 min. 0.6 ml of chloroform/isoamyl alcohol (24:1) was added, the tube was vortex mixed thoroughly and incubated on ice for 30 min. The tube was centrifuged at 12,000 x g for 30 min and the aqueous phase carefully transferred to a fresh microfuge tube without disturbing the interface. An equal volume of isopropanol was added, mixed by inversion and incubated at room temperature for 10 minutes. The tube was centrifuged at 2000 x g for 5 min, the supernatant was removed and the pellet allowed to air dry. The pellet was suspended in 200 µl of 18 MΩ water and incubated at 37° C for 15-30 min. 100 µl of 7.5 M ammonium acetate was added, mixed by inversion and incubated on ice for 1 hour. The tube was centrifuged at 12000 x g for 30 min, the supernatant transferred to a fresh tube and 0.54 volumes of isopropanol were added, mixed by inversion and incubated at room temperature for 10 minutes. The tube was centrifuged at high speed for 5 min, the supernatant was removed and the pellet washed in 500 µl of 70% ethanol. The tube was centrifuged at high speed for 5 min and all the ethanol was removed. The pellet was air dried and suspended in 100 µl of TE (10 mM Tris-HCl pH

7.5; 1mM EDTA) or 18 MΩ water. The DNA was treated with RNase A (1 µl of 1mg/ml stock) before use.

To confirm the transformation of *A. fumigatus* with plasmid DNA containing the *pyrG* gene (ppyrG, pRic3, pMB2 and pMB3) we subjected the purified DNA from transformed fungal colonies to PCR and southern analysis.

PCR was carried out using the following primers:

pyrG1: 5'-gca gag cga ggt atg tag gc-3'(Seq ID No 20);
pyrG2: 5'-aag ccc tcc cgt atc gta gt-3'(Seq ID No 21);
pyrG3: 5'-ata cct gtc cgc ctt tct cc-3'(Seq ID No 22);
pyrG4: 5'-ttt atc cgc ctc cat cca-3'(Seq ID No 23); and
pyrG5: 5'-gcc ttc ctg ttt ttg ctc ac-3'(Seq ID No 24).

All the pyrG primers (pyrG1-pyrG5) were designed to pUC19 sequence which is present in the *pyrG* transformation cassettes. Designing primers to the actual *pyrG* gene would be of no diagnostic use as *pryG*⁻ strains still carry the *pyrG* sequence.

To confirm the transformation of *A. fumigatus* with T-DNA containing the *hph* gene from *A. tumefaciens* we subjected the purified DNA from transformed fungal colonies to PCR and southern analysis.

PCR was carried out using the following primers:

hph6: 5'-cga tgt agg agg gcg tgg at-3'(Seq ID No 1);
hph7: 5'-atc gcc tcg ctc cag tca at-3'(Seq ID No 2);
hph12: 5'-ctt agc cag acg agc ggg tt-3'(Seq ID No 3);
hph13: 5'-caa gac ctg cct gaa acc ga-3'(Seq ID No 4); and
hph14: 5'-tcg tcc atc aca gtt tgc ca-3' (Seq ID No 5).

For southern analysis, approximately 2.5 µg DNA was digested with a restriction enzyme which does not cut within the inserted DNA sequence for 16 hours and separated on a 0.8% agarose TAE gel. DNA was transferred to a Hybond N membrane

by capillary blotting (overnight) and the membrane was pre-hybridized according to the Hybond protocol. Probes specific for the *pyrG* or *hph* gene were digoxigenin (DIG) (Feinberg, A.P. and Vogelstein, B. (1983) Analytical Biochemistry **132**, 6-13) or α^{32}P-labelled PCR products amplified using the primers detailed in this section.

1.9 Isolation of mutated gene sequences

1.9.1 TAIL-PCR

As the T-DNA region of pRic1 does not contain a bacterial origin of replication, plasmid rescue (see section 1.9.2) cannot be used to isolate the mutated gene of interest. In this case we used a method described as thermal asymmetric interlaced PCR (TAIL-PCR: Liu *et al.* (1995) The Plant Journal **8**: 457-463).

Primers were designed to the NPTII gene region of pRic1. This region of DNA is inserted into the mutated gene during transformation and hence acts as a marker for the mutated gene.

The primers used were:

NPT1: 5'-tcc cgc tca gaa gaa ctc gtc aa-3' (Seq ID No 6);

NPT2: 5'-ttg ggt gga gag gct att cgg ct-3'(Seq ID No 7);

NPT3: 5'-tgt tgt gcc cag tca tag ccg aa-3'(Seq ID No 8);

NPT4: 5'-agc cga ata gcc tct cca ccc aa-3'(Seq ID No 9);

NPT5: 5'-cag att att tgg att gag agt ga-3'(Seq ID No 10);

AD1: 5'-ntc ga(g/c) t(a/t)t (g/c)g(a/t) gtt-3'(Seq ID No 11);

AD2: 5'-ngt cga (g/c)(a/t)g ana (a/t)ga a-3'(Seq ID No 12); and

AD3: 5'-(a/t)gt gna g(a/t)a nca nag a-3'(Seq ID No 13).

where n = any base, (x/y)=wobble position.

NPT1 and NPT2 were used to check that the NPTII sequence is present in the transformants. NPT1-NPT3 are nested primers. NPT3 is used in the primary TAIL-PCR, NPT4 in the secondary TAIL-PCR and NPT5 in the tertiary TAIL-PCR. AD1-AD3 are arbitrary degenerate primers. TAIL-PCR cycle settings were as described in the published method (Liu *et al. supra*). PCR fragments isolated from the tertiary

TAIL-PCR by this method were cloned and sequenced by conventional molecular biological techniques.

1.9.2 Plasmid Rescue

As the T-DNA region of pRic2 and pRic3 and plasmids ppyrG, pMB2 and pMB3 all contain a bacterial origin of replication, the technique of plasmid rescue can be used to isolate the mutated gene of interest. Genomic DNA isolated from transformed *A. fumigatus* (section 1.8) was digested to completion with a restriction enzyme that does not cut within the inserted DNA. This digested DNA was then purified and re-ligated with T4 DNA ligase. The random sized, closed circular DNA molecules resulting from this process were then used to transform *E. coli* strain Top10 by electroporation using a Biorad Genepulser (2.1 kV, 200 Ω, 25 μF). Transformed cells were plated on LB agar plates containing ampicillin (100 μg/ml). These plates were incubated at 37 °C until single colonies started to form (these single colonies represent *E. coli* Top10 containing a plasmid with a bacterial origin of replication i.e. from the inserted DNA.). DNA was isolated from these cells and sequenced by conventional molecular biological techniques.

1.9.3 Inverse PCR

In a similar manner to plasmid rescue, genomic DNA isolated from transformed *A. fumigatus* was digested to completion with a restriction enzyme that does not cut within the inserted DNA. This digested DNA was then purified and re-ligated with T4 DNA ligase. Instead of being transformed into *E. coli* as in the plasmid rescue method, the random sized, closed circular DNA molecules resulting from this process were subjected directly to PCR.

This technique may be used when trying to isolate the mutated gene from transformants produced using linearised plasmids which have been cut with a known restriction enzyme. For example, we have designed primers to isolate the mutated gene of interest from transformants produced using pMB3 linearised with *Xba*I.

PCR was carried out using the following primers:

RCpyrG5: 5'-gtt tga ggc gaa ttc tc-3'(Seq ID No 14);

RCpyrG3: 5'-ctc tca gac aga att cgt-3'(Seq ID No 15);

pMB3R: 5'-atc cat cac act ggc g-3'(Seq ID No 16);

T7pro: 5'-taa tac gac tca cta tag gg-3'(Seq ID No 17);

M13-20: 5'-gta aaa cga cgg cca g-3'(Seq ID No 18); and

M13-40 5'-gtt ttc cca gtc acg ac-3'(Seq ID No 19).

1.10 Identification of mutated genes

All DNA sequencing was carried out by external contract (MWG Biotech UK Ltd, Waterside House, Peartree Bridge, Milton Keynes, MK6 3BY). Sequence data obtained was compared to sequences in the public domain databases via BLAST searches (National Centre for Biotechnology Information – http://www.ncbi.nlm.nih.gov).

2. RESULTS

2.1 Haploid transport library

Using the protocols detailed above, the inventors transformed haploid AF300 and AF293 with *A. tumefaciens* (LBA4404 and GV3101) and have approximately 300 (MA1-MA300) transformants frozen down (-80°C) which exhibit growth (at various radial growth rates) on hygromycin (400 µg/ml or higher). One of these transformants (MA205) was taken complete circle to show that *A. fumigatus* could be transformed using the *Agrobacterium*-mediated method and that T-DNA could be integrated at a single gene locus and that locus could be identified by DNA sequencing and homology to known genes examined by the use of bioinformatics.

A. fumigatus transformant MA205 was produced by the transformation of AF300 with *A. tumefaciens* GV3101 containing pRic1. PCR with primers hph6 and hph7 revealed that the hygromycin gene was present in the genomic DNA. Because pRic1 does not contain a bacterial origin of replication (this origin is present in pRic2 and pRic3 by the introduction of pUC18 DNA into the *Hin*dIII site (see section 1.2.2)) which allows the isolation of genic DNA sequences (disrupted gene etc) flanking the inserted T-DNA by plasmid rescue, we used a flanking region isolation method known

as TAIL PCR (see section 1.9.1). The DNA fragment isolated by this method was sequenced by MWG Biotech UK Ltd (see SEQ ID No 27) and was translated in all six reading frames to yield a partial protein sequence of 234 amino acids. Using BLAST (National Centre for Biotechnology Information – http://www.ncbi.nlm.nih.gov) this protein sequence showed strongest sequence homology (41% at the amino acid level) to a hypothetical 35.1 KD protein in the NAM8-GAR1 intergenic region of *Saccharomyces cerevisiae* (sp/P38805/YH08 YEAST and see Figure 12).

Transformation of haploid AF293 using the preferred electroporation method has yielded transformants which are *pyrG*$^+$ by their growth on non selective media (see section 1.7.1) and which have been confirmed as true transformants by diagnostic PCR (see section 1.8). Southern blots of some of these transformants indicating the presence of random, single and/or multiple insertion events is shown in Figure 13 (lanes A-F).

2.2 Diploid general library:

We have produced AF293 and AF300 diploids using the abovedescribed methods (see section 1.4) and have shown that these are in fact true *A. fumigatus* diploids via re-haploidisation with the use of the mitotic inhibitor fluorophenylalanine (FPA) (see section 1.6).

These diploids were used to form mutant banks according to the present invention utilising the transformation methods detailed in section 1.7. Diploid transformants have been isolated which are *pyrG*$^+$ by their growth on non selective media and which have been confirmed as true transformants by diagnostic PCR. A southern blot showing an example of a single integration event in a diploid transformant is shown in Figure 13 (lane G).

CLAIMS

1. A mutant bank of micro-organisms comprising a population of mutant cells in which at least one cell has a mutation that disrupts the activity of at least one gene, said micro-organisms being diploid and being inducible into haploid form.

2. The mutant bank according to claim 1, wherein when the micro-organisms normally occur in a haploid form, they are first induced into the diploid form.

3. The mutant bank according to either claim 1 or 2, wherein a plurality of cells in the population each individually have a mutation which disrupts the activity of at least one gene, so that collectively the said plurality of cells have mutations in a plurality of genes within the genome.

4. The mutant bank according to claim 3, wherein the said plurality of genes makes up 0.001% of the genes within the genome.

5. The mutant bank according to any preceding claim, wherein the micro-organism is a fungus.

6. The mutant bank according to any preceding claim, wherein the micro-organism is a filamentous fungus.

7. The mutant bank according to any preceding claim, wherein the micro-organism is *A. fumigatus*.

8. The mutant bank according to any preceding claim, wherein the micro-organism is AF300 or AF293.

9. The mutant bank according to any of claims 1 to 4, wherein the micro-organism is a yeast.

10. The mutant bank according to claim 9, wherein the micro-organism is *C. glabrata*.

11. The mutant bank according to any preceding claim, wherein the mutations have been randomly induced.

12. The mutant bank according to any preceding claim, wherein the mutant bank is formed by an insertional mutagenesis method, wherein a DNA molecule is inserted into each gene to cause the mutation.

13. The mutant bank according to claim 13, wherein the DNA molecule is a selectable marker.

14. The mutant bank according to claim 13, wherein the selectable marker is a *pyrG* gene.

15. The mutant bank according to claim 14, wherein the *pyrG* gene is from *A. fumigatus*.

16. The mutant bank according to either claim 14 or 15, wherein the *pyrG* gene is harboured on a plasmid.

17. The mutant bank according to claim 16, wherein the *pyrG* harbouring plasmid is an *Aspergillus pyrG* containing plasmid.

18. The mutant bank according to any of claims 1 to 13, wherein the mutations are induced using the Ti plasmid of *Agrobacterium*.

19. The mutant bank according to claim 18, wherein the mutations are induced using the Ti plasmid of *LBA4404 or GV3101*.

20. A method of generating a mutant bank of micro-organisms comprising a population of mutant cells in which at least one cell has a mutation that disrupts the activity of at least one gene, said method comprising the steps of :-

(i) culturing a population of micro-organisms; and

(ii) inducing a mutation in at least one cell of the population which mutation disrupts the activity of at least one gene in the genome of the micro-organism, said micro-organisms being diploid and being inducible into haploid form.

21. A method of generating a mutant bank according to claim 20, wherein when the micro-organisms are normally haploid, they are first induced into the diploid form prior to culturing the said population.

22. A method of generating a mutant bank according to either claim 20 or 21, wherein the micro-organism is a fungus.

23. A method of generating a mutant bank according to any claims 20 or 22, wherein the micro-organism is a filamentous fungus.

24. A method of generating a mutant bank according to any of claims 20 to 23, wherein the micro-organism is *A.fumigatus*.

25. A method of generating a mutant bank according to any of claims 20 to 24, wherein the micro-organism is AF300 or AF293.

26. A method of generating a mutant bank according to either claim 20 or 21, wherein the micro-organism is a yeast.

27. A method of generating a mutant bank according to claim 26, wherein the micro-organism is a *C.glabrata*.

28. A method of generating a mutant bank according to any of claims 20 to 27, wherein the mutations have been randomly induced.

29. A method of generating a mutant bank according to any of claims 20 to 28, wherein the mutant bank is formed by an insertional mutagenesis method, wherein a DNA molecule is inserted into each gene to cause the mutation.

30. A method of generating a mutant bank according to claim 29, wherein the DNA molecule is a selectable marker.

31. A method of generating a mutant bank according to claim 30, wherein the selectable marker is a *pyrG* gene.

32. A method of generating a mutant bank according to claim 31, wherein the *pyrG* gene is from *A. fumigatus*.

33. A method of generating a mutant bank according to either claim 31 or 32, wherein the *pyrG* gene is harboured on a plasmid.

34. The mutant bank according to claim 33, wherein the *pyrG* harbouring plasmid is an *Aspergillus pyrG* containing plasmid.

35. A method of generating a mutant bank according to either claim 33 or 34, wherein the plasmid is introduced into the micro-organism by electroporation.

36. A method of generating a mutant bank according to any of claims 33 to 35, wherein the plasmid is pRic3 as shown in the accompanying drawing, Figure 6.

37. A method of generating a mutant bank according to any of claims 33 to 35, wherein the plasmid is ppyrG as shown in the accompanying drawing, Figure 7.

38. A method of generating a mutant bank according to any of claims 33 to 35, wherein the plasmid is pMB2 as shown in the accompanying drawing, Figure 10.

39. A method of generating a mutant bank according to any of claims 33 to 35, wherein the plasmid is pMB3 as shown in the accompanying drawing, Figure 11.

40. A method of generating a mutant bank according to any of claims 20 to 39, wherein the diploid micro-organism is a mutant for the *pyrG*⁻ phenotype.

41. A method of generating a mutant bank according to claim 40, wherein the diploid mutant micro-organism is generated using the following steps:-

(i) isolating a first haploid mutant comprising a first auxotrophic marker wherein the first auxotrophic marker results in the *pyrG*⁻ phenotype;
(ii) isolating a second haploid mutant using the first haploid mutant isolated in step (i), said second mutant comprising the first *pyrG*⁻ auxotrophic marker and a second auxotrophic marker;
(iii) isolating a third haploid mutant using the first haploid mutant isolated in step (i), said third mutant comprising the first *pyrG*⁻ auxotrophic marker and a third auxotrophic marker; and
(iv) mating the second and third haploid mutants isolated in steps (ii) and (iii) to generate the diploid mutant exhibiting the *pyrG*⁻ phenotype.

42. A method of generating a mutant bank according to claim 41, wherein the second and third auxotrophic markers are selected from *niaD*⁻ or *cnx*⁻.

43. A method of generating a mutant bank according to either claim 41 or 42, wherein step (iv) of the method comprises use of the Frontier method for the production of diploid micro-organisms in *A.fumigatus*.

44. A method of generating a mutant bank according to any of claims 41 to 43, wherein the haploid mutants are incubated at a temperature in the range of 20-36°C.

45. A method of generating a mutant bank according to any of claims 41 to 43, wherein the haploid mutants are incubated at a temperature in the range of 24-33°C.

46. A method of generating a mutant bank according to any of claims 41 to 43, wherein the haploid mutants are incubated at a temperature in the range of 26-30°C.

47. A method of generating a mutant bank according to any of claims 20 to 46, wherein putative mutant cells are identified using a microscope when they are substantially not visible to the naked eye at the earliest stage possible in their development.

48. A method of generating a mutant bank according to claim 47, wherein putative mutant cells are identified at 32 to 48 hours in to their development.

49. A method of identifying genes in a micro-organism which contribute to a chosen phenotype comprising:-

(i) generating a mutant bank of diploid micro-organisms consisting of a population of mutant cells in which at least one cell has a mutation which disrupts the activity of at least one gene;
(ii) exposing the diploid micro-organisms to an agent that induces the micro-organisms into haploid form;
(iii) separating and culturing the haploid micro-organisms as single clones; and
(iv) selecting clones for which the chosen phenotype is altered relative to a wild type micro-organism; and
(v) identifying the mutated gene in each of the selected clones.

50. The method according to claim 49, wherein the mutant bank is as defined in any one of claims 1 to 19 and wherein the mutant bank is generated as is defined in any one of claims 20 to 48.

51. Use of SEQ ID No. 27 for the manufacture of a medicament for treatment of an infection of *A.fumigatus*.

52. A mutant bank of diploid micro-organisms consisting of a population of mutant cells in which each individual cell has a mutation that disrupts the activity of one gene, said population collectively having a mutation in every gene within the genome and wherein the mutant bank may be induced into haploid form.

53. The mutant bank according to claim 52, wherein the micro-organism is *A. fumigatus*.

54. The mutant bank according to claim 53, wherein the micro-organism is AF300 or AF293.

55. The mutant bank according to claim 52, wherein the micro-organism is *C. glabrata*.

56. The mutant bank according to any of claims 52 to 55, wherein the mutations are induced using *Aspergillus pyrG* containing plasmids.

57. The mutant bank according to any of claims 52 to 56, wherein the mutations are induced using the Ti plasmid of *Agrobacterium*.

58. The mutant bank according to claim 57, wherein the mutations are induced using the Ti plasmid of *LBA4404 or GV3101*.

59. A method of identifying genes in a micro-organism which contribute to a chosen phenotype comprising:

(i) generating a mutant bank of diploid micro-organisms consisting of a population of mutant cells in which each individual cell has a mutation which disrupts the activity of one gene, said population collectively having a mutation in every gene within the genome;

(ii) exposing the diploid micro-organisms to an agent that induces the micro-organisms into haploid form;

(iii) separating and culturing the haploid micro-organisms as single clones;

(iv) selecting colonies for which the chosen phenotype is altered relative to a wild type micro-organism; and

(v) identifying the mutated gene in each of the selected clones.

60. The method according to claim 59, wherein the mutant bank is as defined in any one of claims 52 to 59.

61. A method of identifying genes in a micro-organism which contribute to a chosen phenotype essentially as described herein with reference to the Example and accompanying drawings.

FIGURE 1. Plasmid pAN7-1

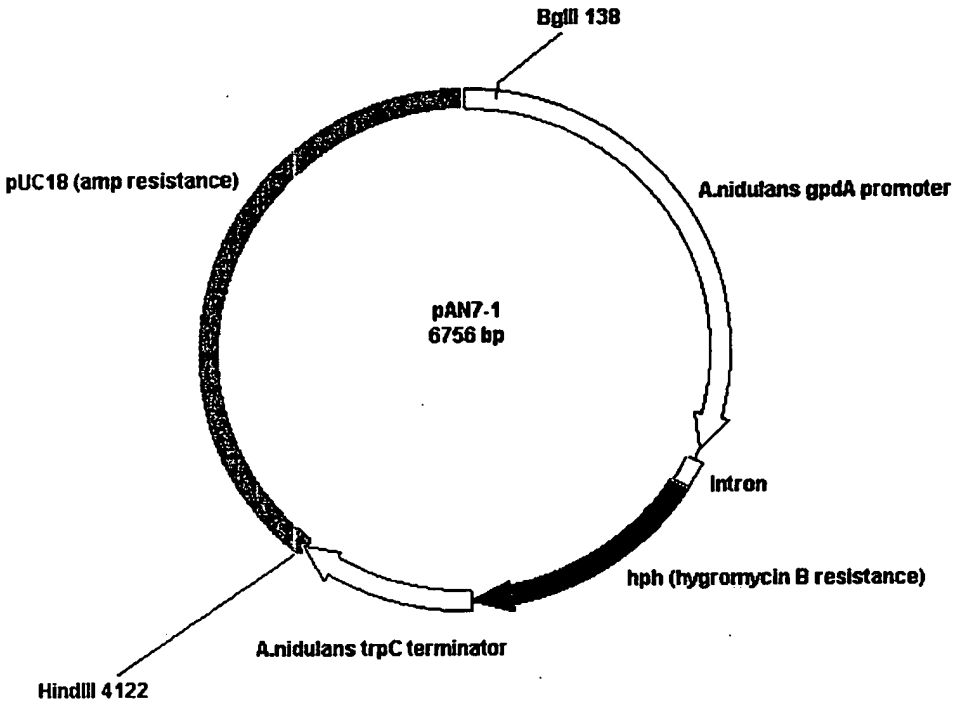


FIGURE 2. Plasmid pRok2

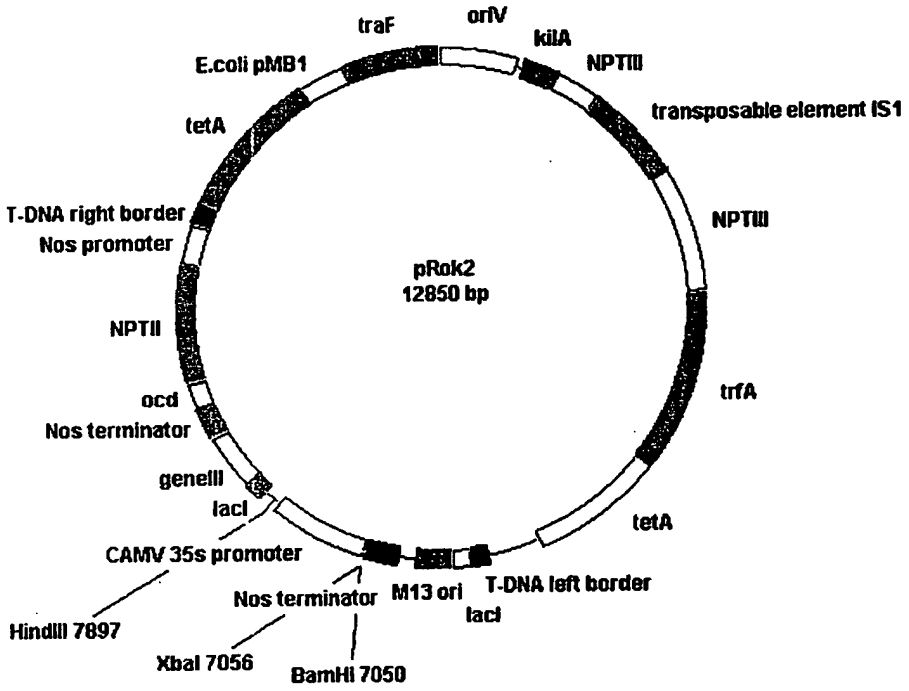


FIGURE 3. Plasmid pRic1

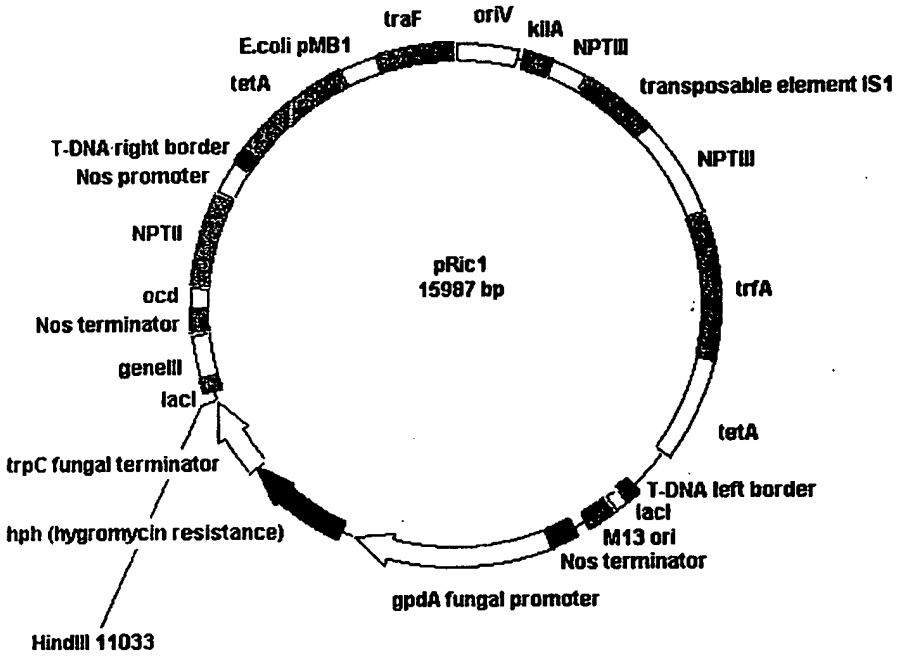


FIGURE 4. Plasmid pUC18

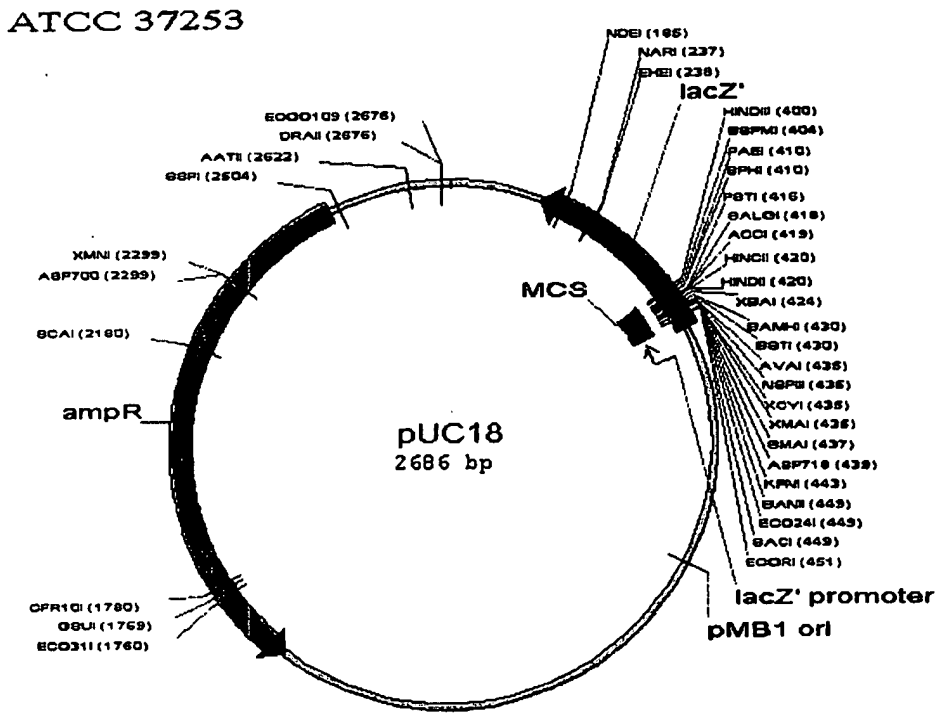


FIGURE 5. Plasmid pRic2

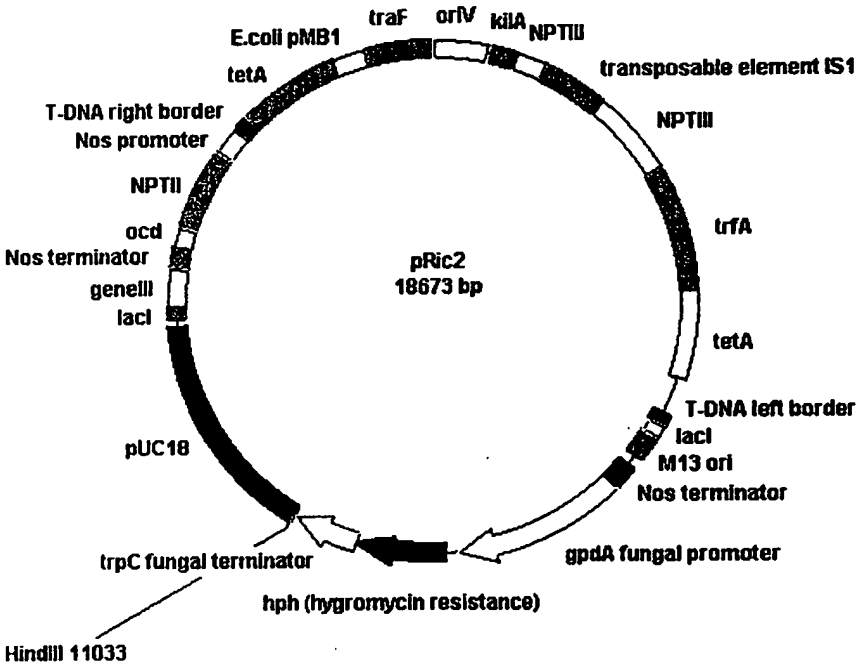


FIGURE 6. Plasmid pRic3

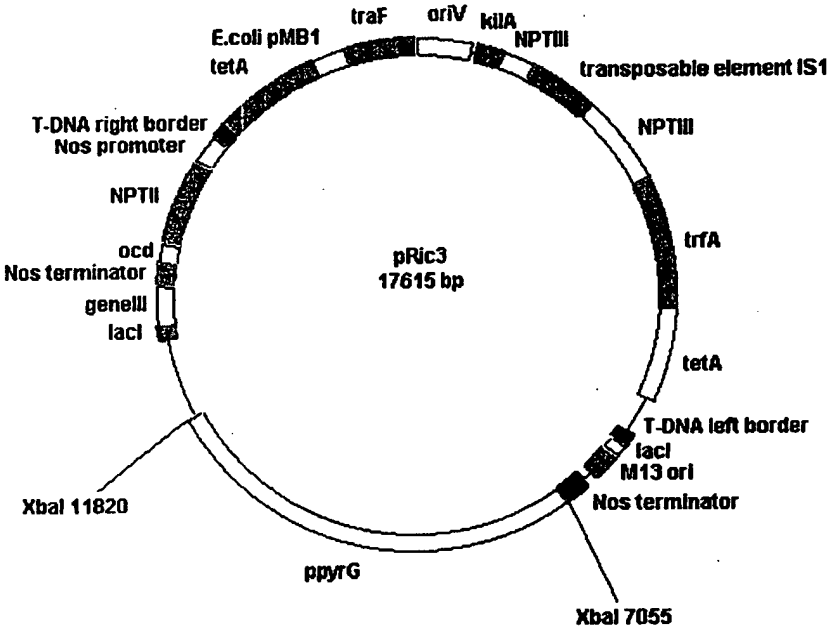


FIGURE 7. Plasmid ppyrG

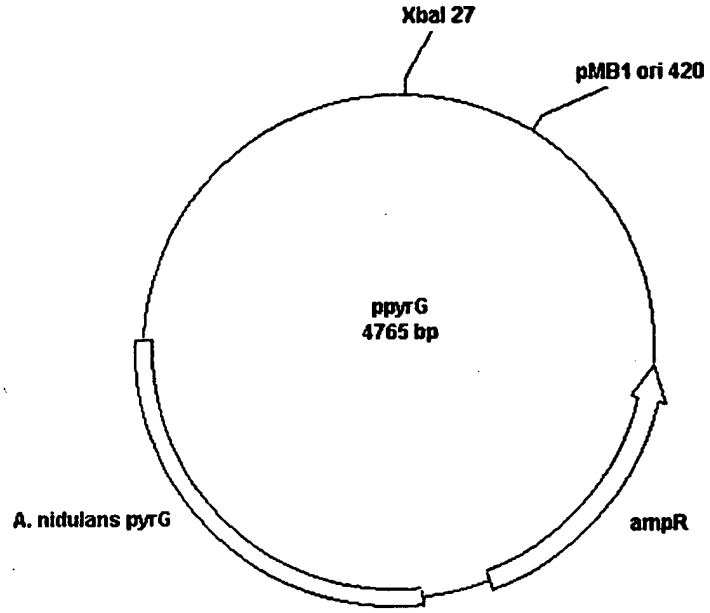


FIGURE 8. Plasmid pCR2.1

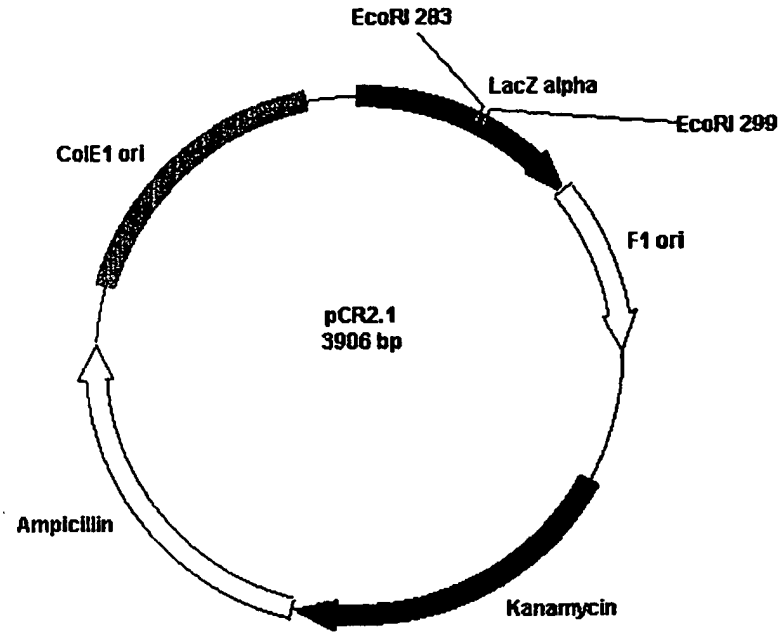


FIGURE 9. Plasmid pTT

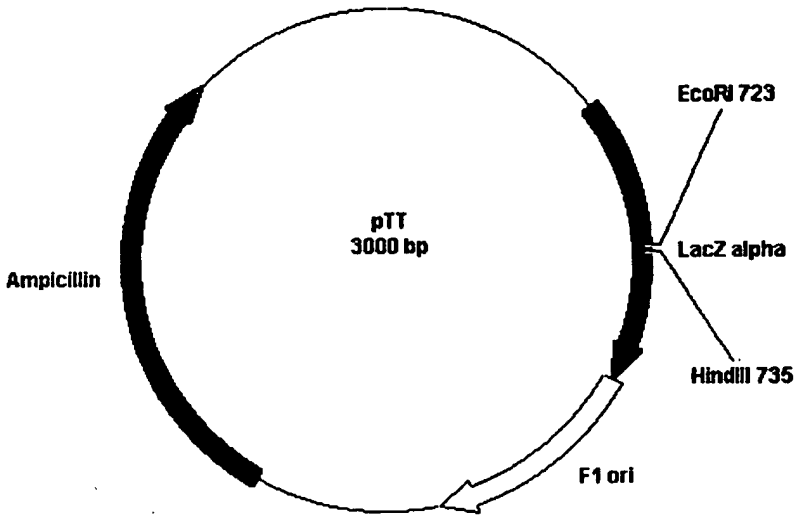


FIGURE 10. Plasmid pMB2

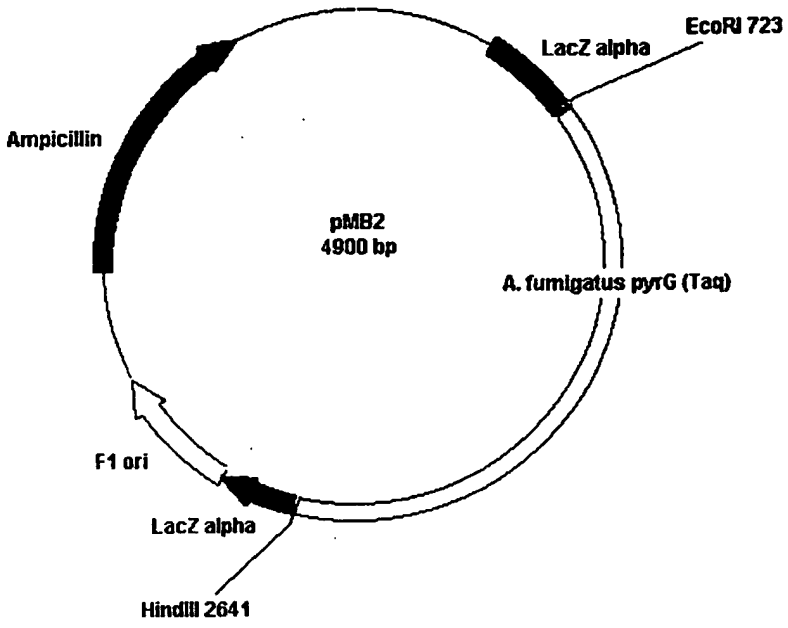


FIGURE 11. Plasmid pMB3

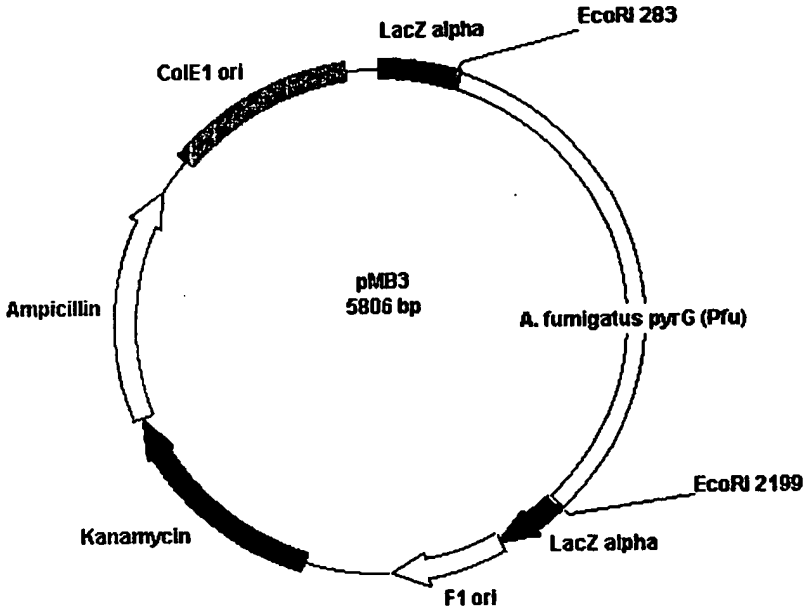


FIGURE 12. Identification of mutated gene in strain MA205

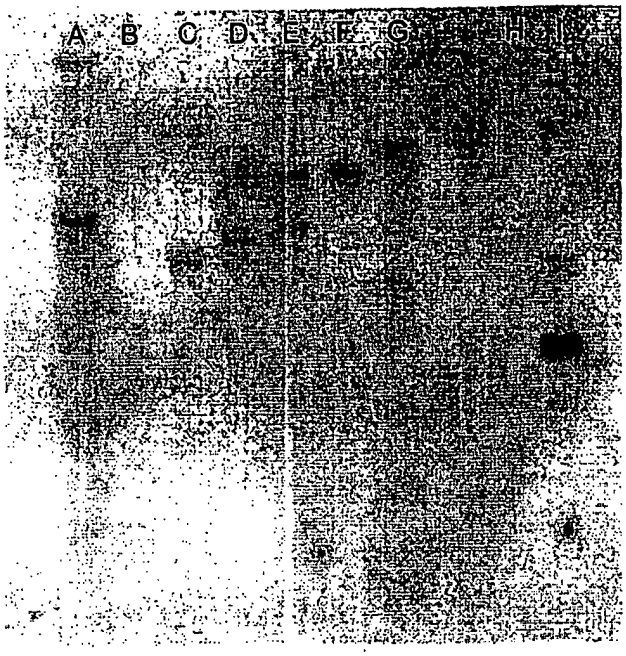
BLASTX search identifies two yeast proteins as nearest matches:

Sequences producing significant alignments:	Score (bits)	E Value
sp|P38805|YH08_YEAST HYPOTHETICAL 35.1 KD PROTEIN IN NAM8-GAR1 ...	194	5e-49
sp|O14180|YDS4_SCHPO HYPOTHETICAL 35.8 KD PROTEIN C4F8.04 IN CH...	173	1e-42
gi|3695379 (AF096370) contains similarity to a C. elegans hypot...	143	8e-34
emb|CAA93858| (Z70034) similarity to 35.1KD hypothetical yeast ...	113	1e-24
gb|AAD14602| (AF092910) stage specific peptide 24 [Trypanosoma ...	81	8e-15
emb|CAB11643| (Z98974) hypothetical protein [Schizosaccharomyce...	76	2e-13
emb|CAB05841| (Z83246) predicted using Genefinder; cDNA EST EMB...	75	4e-13
sp|P53941|YNH5_YEAST HYPOTHETICAL 33.5 KD PROTEIN IN MKS1-MSK1 ...	74	8e-13
emb|CAB11063| (Z98531) hypothetical protein [Schizosaccharomyce...	33	1.7
gi|2650581 (AE001102) ATP-dependent RNA helicase, putative [Arc...	33	2.2
emb|CAA74646| (Y14274) putative serine/threonine protein kinase...	32	4.9

sp|P38805|YH08_YEAST HYPOTHETICAL 35.1 KD PROTEIN IN NAM8-GAR1 INTERGENIC REGION
>gi|626640|pir||S46718 hypothetical protein YHR088w - yeast (Saccharomyces cerevisiae) >gi|487932 (U00060) Yhr088wp [Saccharomyces cerevisiae]
Length = 295

Score = 194 bits (488), Expect = 5e-49
Identities = 97/234 (41%), Positives = 146/234 (61%), Gaps = 29/234 (12%)

FIGURE 13. Southern Blot showing random integration events



Genomic DNA digested with EcoRI. Probe was pUC18 labelled using the DIG method.

Lane

A. Single integration. Uncut ppyrG plasmid. Electroporation of conidia. AF293 haploid *pyrG⁻* strain.

B. Single integration. Uncut ppyrG plasmid. Electroporation of conidia. AF293 haploid *pyrG⁻* strain.

C. Double integration. Uncut ppyrG plasmid. Electroporation of conidia. AF293 haploid *pyrG⁻* strain.

D. Double integration. Uncut ppyrG plasmid. Electroporation of conidia. AF293 haploid *pyrG⁻* strain.

E. Double integration. Uncut ppyrG plasmid. Electroporation of conidia. AF293 haploid *pyrG⁻* strain.

F. Single integration. Uncut ppyrG plasmid. Electroporation of conidia. AF293 haploid *pyrG⁻* strain.

G. Single integration. XbaI cut ppyrG plasmid. Protoplast transformation. AF293 diploid *pyrG⁻/ pyrG⁻/niaD⁻/cnx⁻* strain.

H. Negative control

I. Positive control

SEQUENCE LISTING

<110> F2G Limited

<120> MUTANT BANK

<130> AH/DPW/U514

<140> GB0008748.6
<141> 2000-04-11

<160> 27

<170> PatentIn version 3.0

<210> 1
<211> 20
<212> DNA
<213> Escherichia coli

<400> 1
cgatgtagga gggcgtggat 20

<210> 2
<211> 20
<212> DNA
<213> Escherichia coli

<400> 2
atcgcctcgc tccagtcaat 20

<210> 3
<211> 20
<212> DNA
<213> Escherichia coli

<400> 3
cttagccaga cgagcgggtt 20

<210> 4
<211> 20
<212> DNA
<213> Escherichia coli

<400> 4
caagacctgc ctgaaaccga 20

<210> 5
<211> 20
<212> DNA
<213> Escherichia coli

<400> 5
tcgtccatca cagtttgcca 20

<210> 6
<211> 23
<212> DNA
<213> Escherichia coli

<400> 6
tcccgctcag aagaactcgt caa 23

<210> 7
<211> 23
<212> DNA
<213> Escherichia coli

<400> 7
ttgggtggag aggctattcg gct 23

<210> 8
<211> 23
<212> DNA
<213> Escherichia coli

<400> 8
tgttgtgccc agtcatagcc gaa 23

<210> 9
<211> 23
<212> DNA
<213> Escherichia coli

<400> 9
agccgaatag cctctccacc caa 23

<210> 10
<211> 23
<212> DNA
<213> Agrobacterium tumefaciens

<400> 10
cagattattt ggattgagag tga 23

<210> 11
<211> 15
<212> DNA
<213> arbitrary degenerate

<220>
<221> misc_feature

<222> (1)..(1)
<223> a, g, c or t

<220>
<221> misc_feature
<222> (6)..(6)
<223> g or c

<220>
<221> misc_feature
<222> (8)..(8)
<223> a or t

<220>
<221> misc_feature
<222> (10)..(10)
<223> g or c

<220>
<221> misc_feature
<222> (12)..(12)
<223> a or t

<400> 11
ntcgantntn gngtt 15

<210> 12
<211> 16
<212> DNA
<213> arbitrary degenerate

<220>
<221> misc_feature
<222> (1)..(1)
<223> a, g, c or t

<220>
<221> misc_feature
<222> (7)..(7)
<223> g or c

<220>
<221> misc_feature
<222> (8)..(8)
<223> a or t

<220>

<221> misc_feature
<222> (11)..(11)
<223> a, g, c or t

<220>
<221> misc_feature
<222> (13)..(13)
<223> a or t

<400> 12
ngtcgannga nangaa 16

<210> 13
<211> 16
<212> DNA
<213> arbitrary degenerate

<220>
<221> misc_feature
<222> (1)..(1)
<223> a or t

<220>
<221> misc_feature
<222> (5)..(5)
<223> a, g, c or t

<220>
<221> misc_feature
<222> (8)..(8)
<223> a or t

<220>
<221> misc_feature
<222> (10)..(10)
<223> a, g, c or t

<220>
<221> misc_feature
<222> (13)..(13)
<223> a, g, c or t

<400> 13
ngtgnagnan canaga 16

<210> 14
<211> 17

<212> DNA
<213> Aspergillus fumigatus

<400> 14
gtttgaggcg aattctc 17

<210> 15
<211> 18
<212> DNA
<213> Aspergillus fumigatus

<400> 15
ctctcagaca gaattcgt 18

<210> 16
<211> 16
<212> DNA
<213> Escherichia coli

<400> 16
atccatcaca ctggcg 16

<210> 17
<211> 20
<212> DNA
<213> Escherichia coli

<400> 17
taatacgact cactataggg 20

<210> 18
<211> 16
<212> DNA
<213> Escherichia coli

<400> 18
gtaaaacgac ggccag 16

<210> 19
<211> 17
<212> DNA
<213> Escherichia coli

<400> 19
gttttcccag tcacgac 17

<210> 20
<211> 20
<212> DNA
<213> Escherichia coli

<400> 20
gcagagcgag gtatgtaggc 20

<210> 21
<211> 20
<212> DNA
<213> Escherichia coli

<400> 21
aagccctccc gtatcgtagt 20

<210> 22
<211> 20
<212> DNA
<213> Escherichia coli

<400> 22
atacctgtcc gcctttctcc 20

<210> 23
<211> 18
<212> DNA
<213> Escherichia coli

<400> 23
tttatccgcc tccatcca 18

<210> 24
<211> 20
<212> DNA
<213> Escherichia coli

<400> 24
gccttcctgt ttttgctcac 20

<210> 25
<211> 24
<212> DNA
<213> Aspergillus fumigatus

<400> 25
ctacctcgag aattcgcctc aaac 24

<210> 26
<211> 22
<212> DNA
<213> Aspergillus fumigatus

<400> 26

ggcgacgaat tctgtctgag ag 22

<210> 27
<211> 1060
<212> DNA
<213> Aspergillus fumigatus

<400> 27
ctcgaggtcg acggtatcga taagcttgat tagccgaata gctctccacc caaaatcctc 60

ttcaccacct cgctcaactc taccctccat aacgaggcca gatcccctac cgatctgttt 120

cctaacagcg tttatatccg tcgtacggcc catcgctatt cccacaagtt ctccatccgc 180

gagattgctg ccttcgcgtc gaatcggaac tataccagcg tcgtggttct ccaggaggat 240

cagaaacgac cttccggtct gacgatcgtc cacctacctc agggtctcac cttccatttt 300

actattagca gatggataga aggcaagaag cttccgggac atggaaatgc caccgagcac 360

cggccggaac taatcctcaa caacttccgg actcctcctg gtctacttac ggcccatctc 420

ttccggagca tgttcccacc acaaccggac ctagaaggac gtcaagttgt cgcaattcac 480

aaccaaagag actacctatt tctccggcgt catcgttatg tgttccgtga gaagcgcgag 540

actgaaaaga gtgtcgttgg tacagacggc aaagagatca agggtgtgga aggcattcgc 600

gccggccttc aagagcttgg tcccgcaata aacttcaagc tgcggcgagt tgacaaagga 660

attcagcgtg ctagcggaca ggagtgggag tggaaaggag gcatggaaaa acagcgtact 720

aagttccagc tgtagcatgc tctcatcccc ttttatacct tttgttatgt gtctgattgc 780

atcattgggt ttgcatggag tttagctttc aagaatgggt ctttggatgg ctttttcacg 840

tgcggttata gtttcttcat gactgtccag atccttatac catctgtaga taatgtactg 900

aaaaccacta ttctatcaaa gtggtgtttg ggccaaatgg tggcatctca tctactctca 960

aatgcacaca tgcgattgat caatagataa gatgctcaaa tacaacaatc tgtaatccat 1020

gttttgattc ccgggggggta cccgaaatcg aattcctgca 1060

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 01/01626

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N1/16 C12N15/80 A61K48/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, BIOSIS, CHEM ABS Data, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 99 15644 A (CARDINAL GUY ;UNIV LAVAL (CA); LEVESQUE ROGER C (CA); SANSCHAGRIN) 1 April 1999 (1999-04-01) page 20, paragraph 1 page 23, line 25 - line 28	1-61

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
1 August 2001	09/08/2001

Name and mailing address of the ISA
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer
Mata Vicente, T.

1

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

nformation on patent family members

Int · ·onal Application No
PCT/GB 01/01626

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9915644 A	01-04-1999	AU 9247698 A	12-04-1999

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 March 2001 (01.03.2001)

PCT

(10) International Publication Number
WO 01/14533 A2

(51) International Patent Classification[7]: C12N 9/00

(21) International Application Number: PCT/GB00/03100

(22) International Filing Date: 15 August 2000 (15.08.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
9919766.7 21 August 1999 (21.08.1999) GB

(71) Applicant *(for all designated States except MG, US)*: ASTRAZENECA AB [SE/SE]; S-151 85 Sodertalje (SE).

(71) Applicant *(for MG only)*: ASTRAZENECA UK LIMITED [GB/GB]; 15 Stanhope Gate, London W1Y 6LN (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: ROSAMOND, John, David, Charles [GB/US]; 35 Gatehouse Drive, Waltham, MA 02451 (US). SCHNELL, Norbert, Friedemann [DE/US]; 35 Gatehouse Drive, Waltham, MA 02451 (US).

(74) Agent: GAINEY, Laurence, David, Scott; Astrazeneca, Global Intellectual Property, PO Box 272, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4GR (GB).

(81) Designated States *(national)*: JP, MG, US.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/14533 A2

(54) Title: PROTEIN

(57) Abstract: This invention relates to newly identified polynucleotides, polypeptides encoded by these polynucleotides, to the production of such polynucleotides and polypeptides, and to the uses of such polynucleotides and polypeptides. More specifically, the invention relates to the phosphomevalonate kinase (PMK) gene (ERG8 gene) from *Candida Albicans* (*C. albicans*), to methods for its expression yielding phosphomevalonate kinase protein, to novel hybrid organisms for use in such expression methods, to methods for purification of the protein, to methods and tools for diagnosing *C. albicans* infection and to assays for identifying inhibitors of the enzyme which inhibitors have potential as anti-fungal agents.

PROTEIN

This invention relates to newly identified polynucleotides, polypeptides encoded by these polynucleotides, to the production of such polynucleotides and polypeptides, and to the uses of such polynucleotides and polypeptides. More specifically, the invention relates to the phosphomevalonate kinase (PMK) gene (ERG8 gene) from *Candida albicans* (*C. albicans*), to methods for its expression yielding phosphomevalonate kinase protein, to novel hybrid organisms for use in such expression methods, to methods for purification of the protein, to methods and tools for diagnosing *C. albicans* infection and to assays for identifying inhibitors of the enzyme which inhibitors have potential as anti-fungal agents.

C. albicans is an important human fungal pathogen and the most prominent target organism for antifungal research. PMK is an enzyme required for the biosynthesis of isoprene subunits that are used as precursors in the synthesis of sterols, dolichols and ubiquinones. As PMK is an essential biosynthetic enzyme, inhibitors of PMK should find use as antifungal agents. All species synthesise a protein with PMK activity however, across species the enzymes differ considerably in their amino acid sequence. Because of selectivity problems (for example fungal versus human) it is extremely important to optimise potential inhibitors specifically against the fungal target enzymes (i.e. *C. albicans* or *Aspergillus fumigatus*) and not against the human enzyme. Such cross-fungal-species inhibitors possess broad specificity. Alternatively, it may be desirable to use an inhibitor which is more selective, for example, one that inhibits *C. albicans* PMK but not a homologous but non-identical fungal PMK protein such as from *Saccharomyces cerevisiae* (*S. cerevisiae*).

In view of the increased incidence of fungal resistance to existing anti-fungal agents and fuelled by the growing number of fungal infections particularly in people with immunodeficiency disorders, organ transplants and cancer, there is a need for new means of identifying potential anti-fungal agents.

We have now successfully cloned the ERG8 gene from *C. albicans* (hereinafter referred to as ERG8 gene) and determined its full length nucleotide sequence and corresponding (PMK) polypeptide sequence (hereinafter referred to as ERG8 protein) as set out in Figure 1 and SEQ ID No. 7 of this application respectively. The coding DNA sequence (SEQ ID NO. 6) of the *C. albicans* ERG8 gene isolated is 1299 nucleotides in length and the corresponding protein sequence is 433 amino acids in length (SEQ ID NO. 7). The protein exhibits approximately 45% homology with the corresponding protein from *S. cerevisiae* and

only about 10% homology to that of the human protein equivalent. Homology as used herein, takes the definition known to and routinely used by molecular biologists. It refers to the sequence identity between two sequences as assessed by best-fit computer alignment analysis using suitable software such as Blast, Blast2, NCBI Blast2, WashU Blast2, FastA, Fasta3 and PILEUP, using a scoring matrix such as Blosum 62. Such software packages endeavour to closely approximate the "gold-standard" alignment algorithm of Smith-Waterman. Thus, the preferred software/search engine programme for use in assessing the percent identity or similarity, i.e how two primary polypeptide sequences line up is Smith-Waterman. Identity refers to direct matches, similarity allows for conservative substitutions.

According to a first aspect of the invention there is provided an isolated or purified polypeptide which is ERG8 protein, as well as variants thereof. The preferred polypeptide sequence is that as set out in SEQ ID NO. 7. The complete *C. albicans* phosphomevalonate kinase enzyme polypeptide has the amino acid sequence as depicted in SEQ ID No. 7 herein. The polypeptides of the present invention include the polypeptide of SEQ ID No. 7 as well as polypeptides which have in increasing order of preference, at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, and 99% identity to the polypeptide whose amino acid sequence is depicted in SEQ ID NO. 7.

As used herein, the term "isolated" refers to molecules, either nucleic acid or amino acid sequences, that are removed from their natural environment and purified or separated from at least one other component with which they are naturally associated. Also encompassed by this term are molecules that are artificially synthesised and purified away from their synthesis materials. Thus, a polynucleotide is said to be isolated when it is substantially separated from other contaminant polynucleotides or nucleotides.

Although the natural polypeptide of SEQ ID NO. 7 and a variant polypeptide may only possess for example 80% identity, they are actually likely to possess a higher degree of similarity, depending on the number of dissimilar codons that are conservative changes. Similarity between two sequences includes direct matches as well as conserved amino acid substitutes which possess similar structural or chemical properties, e.g. similar charge. Examples of conservative changes (conserved amino acid substitutes) are *inter alia*: alanine to glycine, isoleucine, valine or leucine; tyrosine to phenylalanine or tryptophan; and lysine to arginine or histidine.

Suitable conservative substitutions of amino acids are known to those of skill in this art and may be made without altering the biological activity of the resulting polypeptide,

regardless of the chosen method of synthesis. The phrase "conservative substitution" includes the use of a chemically derivatized residue in place of a non-derivatized residue provided that such polypeptide displays the desired binding activity. D-isomers as well as other known derivatives may also be substituted for the naturally occurring amino acids. *See, e.g.*, U.S. Patent No. 5,652,369, *Amino Acid Derivatives*, issued July 29, 1997. Substitutions are preferably, although not exclusively, made in accordance with those set forth in TABLE 1 as follows:

TABLE 1

Original residue	Example conservative substitution
Ala (A)	Gly; Ser; Val; Leu; Ile; Pro
Arg (R)	Lys; His; Gln; Asn
Asn (N)	Gln; His; Lys; Arg
Asp (D)	Glu
Cys (C)	Ser
Gln (Q)	Asn
Glu (E)	Asp
Gly (G)	Ala; Pro
His (H)	Asn; Gln; Arg; Lys
Ile (I)	Leu; Val; Met; Ala; Phe
Leu (L)	Ile; Val; Met; Ala; Phe
Lys (K)	Arg; Gln; His; Asn
Met (M)	Leu; Tyr; Ile; Phe
Phe (F)	Met; Leu; Tyr; Val; Ile; Ala
Pro (P)	Ala; Gly
Ser (S)	Thr
Thr (T)	Ser
Trp (W)	Tyr; Phe
Tyr (Y)	Trp; Phe; Thr; Ser
Val (V)	Ile; Leu; Met; Phe; Ala

The nucleotide sequences of the present invention may also be engineered in order to alter a coding sequence for a variety of reasons, including but not limited to, alterations which

modify the cloning, processing and/or expression of the gene product. For example, mutations may be introduced using techniques which are well known in the art, eg, site-directed mutagenesis to insert new restriction sites, to alter glycosylation patterns, to change codon preference, etc.

Included within the scope of the present invention are alleles of the ERG8 molecule of the present invention. As used herein, an "allele" or "allelic sequence" is an alternative form of the kinase molecule described herein. Alleles result from nucleic acid mutations and mRNA splice-variants which produce polypeptides whose structure or function may or may not be altered. Any given gene may have none, one or many allelic forms. Common mutational changes which give rise to alleles are generally ascribed to natural deletions, additions or substitutions of amino acids. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

Thus, according to a preferred embodiment there is provided an isolated polypeptide comprising the sequence depicted in SEQ ID No. 7 or a sequence possessing at least 80% similarity thereto. More preferred embodiments are those that have in increasing order of preference at least 85, 90, 95, 96, 97, 98 and 99% similarity to the sequence depicted in SEQ ID No. 7. Functional biologically active variants are preferred.

Fragments of such polypeptides comprising at least 15, preferably at least 30 and more preferably at least 50 contiguous amino acids are also encompassed by the present invention. Such fragments may be used as intermediates to generate longer polypeptide fragments including preferably, the full-length polypeptide sequence as depicted in SEQ ID No. 7, or a functional variant thereof. Such polypeptide fragments may also be used to raise antibodies against or specific for parts of the ERG8 protein.

The invention also relates to variant polypeptide sequences encoded by nucleic acid capable of hybridising with nucleic acid coding for the natural polypeptide (SEQ ID No. 6, or its complementary antisense strand)(or would do so but for the degeneracy of the genetic code), for example under stringent conditions (such as at 35°C to 65°C in a salt solution of approximately 0.9M). Such hybridisable polynucleotides are also part of the invention. The present invention particularly relates to polynucleotides which hybridise to the ERG8 polynucleotide sequence depicted in SEQ ID NO. 6, its complementary sequence, or fragment thereof, under stringent conditions. As used herein, stringent conditions are those conditions which enable sequences that possess at least 80%, preferably at least 90% and more preferably

at least 95% sequence identity to hybridise together. Thus, nucleic acids which can selectively hybridise to the nucleic acid of SEQ ID No. 6, or the complementary antisense strand thereof, include nucleic acids which have at least 80%, preferably at least 90%, more preferably at least 95%, still more preferably at least 98% sequence identity and most preferably 100%, over at least a portion of the nucleic acid encoding the ERG8 gene disclosed herein. Selectively hybridise means that the molecule must be capable of specifically hybridising to the nucleic acid sequence of SEQ ID No. 6 or its complement, to the exclusion of other naturally occurring sequences. As well as full-length gene sequences, smaller nucleic acid fragments for example oligonucleotide primers which can be used to amplify the ERG8 gene using any of the well known amplification systems such as polymerase chain reaction (PCR), or fragments that can be used as diagnostic probes to identify corresponding nucleic acid sequences are also part of this invention. The invention thus includes polynucleotides of shorter length than the full length ERG8 gene sequence depicted in SEQ ID No. 6, that are capable of specifically hybridising to the nucleic acid encoding the C. albicans ERG8 gene described herein. Such polynucleotides may be at least 10 nucleotides in length, preferably at least 15, more preferably at least 20 and most preferably at least 30 nucleotides in length and may be of any size up to and including the full length ERG8 nucleotide sequence. The presence of mismatch nucleotides in the hybridisation polynucleotides is not detrimental to the utility of such polynucleotides provided that they are capable of selectively hybridising to the target ERG8 nucleotide sequence.

An example of a suitable hybridisation solution when a nucleic acid is immobilised on a nylon membrane and the probe nucleic acid is greater than 500 bases or base pairs is: 6 x SSC (saline sodium citrate), 0.5% SDS (sodium dodecyl sulphate), 100µg/ml denatured, sonicated salmon sperm DNA. The hybridisation being performed at 68°C for at least 1 hour and the filters then washed at 68°C in 1 x SSC, or for higher stringency, 0.1 x SSC/0.1% SDS.

An example of a suitable hybridisation solution when a nucleic acid is immobilised on a nylon membrane and the probe is an oligonucleotide of between 12 and 50 bases is: 3M trimethylammonium chloride (TMACl), 0.01M sodium phosphate (pH 6.8), 1mM EDTA (pH 7.6) , 0.5% SDS,100µg/ml denatured, sonicated salmon sperm DNA and 0.1 dried skimmed milk. The optimal hybridisation temperature (Tm) is usually chosen to be 5°C below the Ti of the hybrid chain. Ti is the irreversible melting temperature of the hybrid formed between the

probe and its target. If there are any mismatches between the probe and the target, the Tm will be lower. As a general guide, the recommended hybridisation temperature for 17-mers in 3M TMACl is 48-50°C; for 19-mers,it is 55-57°C; and for 20-mers, it is 58-66°C.

A suitable hybridisation protocol is described in Example 5 herein, however, operable variations to this method will be apparent to the person skilled in the art.

As used herein, the term 'variant' includes naturally occurring allelic variants as well as non-naturally occurring variants, fragments and analogs of the sequences depicted in SEQ ID NOs. 6 or 7. Such variants include C- or N-truncated variants, deletion variants, substitution variants as well as addition and insertion variants. The term 'analog' refers to proproteins which can be activated by cleavage of the proprotein portion to release the biologically active polypeptide or protein. The term 'derivative' refers to a polypeptide encoded by a chemically modified ERG8 gene, for example one wherein hydrogen has been replaced by an acyl or amino group, as well as polypeptides possessing one or more non-natural amino acids. When referring to a polypeptide or protein sequence, a functional variant is one that has retained at least some PMK enzymatic activity. The variant polypeptides of the present invention may comprise internal, but preferably, terminal flanking sequences (fusion proteins) to facilitate protein purification. Such 'additional domain' sequences (Flag sequences) may comprise for example, metal chelating peptides such as histidine-tryptophan modules (including 6-his tags) that allow purification of the polypeptide on immobilised metals, protein A domains that allow purification on immobilised immunoglobulin, or peptide domains that allow purification on immobilised antibodies specific for the peptide. Other suitable 'additional purification domains' will be known to the person skilled in the art.

According to a preferred embodiment of the invention the native ERG8 polypeptide sequence (having the sequence as depicted in SEQ ID No. 7) is fused at its amino terminus to six histidine residues which serve to enable the polypeptide, once expressed from the host cell, to be isolated and purified by affinity chromatography using a Ni-chelate resin.

A flanking purification domain may be separated from the ERG8 polypeptide by a cleavage sequence such as that recognised by thrombin or Factor Xa so as to facilitate release of the polypeptide from the flanking sequence which may or may not be attached to an immobilised support. Alternatively, cyanogen bromide which cleaves at methionine residues can be employed to release the desired polypeptide from its flanking sequence.

The polypeptides of the invention can be synthesised chemically. For example, by the Merryfield technique (J. Amer. Chem. Soc. 85:2149-2154, 1968). Numerous automated polypeptide synthesisers, such as Applied Biosystems 431A Peptide Synthesizer also now exist. Alternatively, and preferably, the polypeptides of the invention are produced from a nucleotide sequence encoding the polypeptide using recombinant expression technology.

In a further aspect of the invention there are provided isolated polynucleotides (including genomic DNA, genomic RNA, cDNA and mRNA; double stranded as well as +ve and -ve strands) which encode the polypeptides of the invention. Single stranded DNA molecules of all or part of the ERG8 gene either +ve or -ve strand, find use inter alia, as hybridisation probes or PCR amplification primers. The sense strand of the complete gene sequence of native ERG8 is depicted in Figure 1 (SEQ ID No. 5) hereinafter. It will be appreciated that a polynucleotide of the invention may comprise any of the degenerate codes for a particular amino acid, including the use of rare codons. Indeed, when producing the polypeptide by recombinant expression in heterologous host strains, it may be desirable to adopt the codon usage (preference) of the host organism (Murray. N.A.R. 17:477-508, 1989).

Thus, according to a further aspect invention there is provided an isolated polynucleotide comprising nucleic acid encoding the amino acid sequence depicted in SEQ ID No. 7 or a variant thereof, such as one possessing at least 80% identity thereto.

The invention further comprises convenient fragments of any one of the above polynucleotide/nucleic acid sequences. Convenient fragments may be defined by restriction endonuclease digests of nucleic acid comprising the ERG8 gene sequence. Such fragments are useful inter alia, for expressing short polypeptides fragments of ERG8 protein of the invention as well as for use as hybridisation probes. The present invention also provides a polynucleotide probe comprising any one of the above sequences or fragments together with a convenient label or marker, preferably a non-radioactive label or marker. Following procedures well known in the art, the probes can be used to identify and isolate not only corresponding nucleic acid sequences (i.e *C. albicans* ERG8 gene sequences) but, if sufficiently homologous, can also be used to identify the analogous gene from other organisms using techniques well known to the person skilled in the art. Such sequences may be comprised in libraries, such as genomic or cDNA libraries. The present invention also provides RNA transcripts corresponding to any of the above *C. albicans* ERG8 sequences or fragments. RNA transcripts can be used to prepare a polypeptide of the invention by *in vitro* translation techniques according to known methods (Sambrook *et al.* "Molecular Cloning- A

Laboratory Manual, second edition 1989"). The invention further comprises full-length or fragment lengths of ERG8 gene (coding sequence) flanked by non-coding sequence which may include natural or non-natural sequence containing restriction enzyme recognition sequence motifs. The incorporation of suitable restriction enzyme recognition sites either side of the ERG8 coding region, or indeed any polynucleotide sequence from ERG8, facilitates cloning of the ERG8 gene or polynucleotide sequence into a suitable vector. A suitable polynucleotide comprises a full length *C. albicans* ERG8 gene (encoding the polypeptide that starts with methionine at position 1 and terminates with the leucine that precedes the stop codon TAA at position 1299 of Figure 1) flanked by unique HindIII (5'-end)-XhoI (3'-end) restriction sites. Examples of oligonucleotide primers which are suitable for use in PCR amplification of ERG8, and which incorporate useful restriction enzyme sites to facilitate cloning, are disclosed as SEQ ID Nos. 10 and 11. Nucleotide changes or mutations may be introduced into a polynucleotide sequence by *de novo* polynucleotide synthesis, by site directed mutagenesis using appropriately designed oligonucleotide primers or by any other convenient means know to the person skilled in the art.

For expression purposes, it may be advantageous to engineer a restriction site at the 5'-end which is also capable of reconstituting the native amino-terminal methionine of the protein. The cleavage recognition sequence for the NcoI restriction enzyme not only includes a sequence that codes for methionine, but also one that is capable of retaining a functional Kozak consensus sequence, enabling the ERG8 gene to be cloned at the 3'-end of a suitable promoter element in an expression vector.

The polynucleotides can be synthesised chemically, or isolated by one of several approaches known to the person skilled in the art such as polymerase chain reaction (PCR) or ligase chain reaction (LCR) or by cloning from a genomic or cDNA library.

Once isolated or synthesised, a variety of expression vector/host systems may be used to express ERG8 coding sequences. These include, but are not limited to microorganisms such as bacteria expressed with plasmids, cosmids or bacteriophage; yeasts tranformed with expression vectors; insect cell systems transfected with baculovirus expression systems; plant cell systems transfected with plant virus expression systems, such as cauliflower mosaic virus; or mammalian cell systems (for example those transfected with adenoviral vectors); selection of the most appropriate system is a matter of choice.

Expression vectors usually include an origin of replication, a promoter, a translation initiation site, optionally a signal peptide, a polyadenylation site, and a transcription

termination site. These vectors also usually contain one or more antibiotic resistance marker gene(s) for selection. As noted above, suitable expression vectors may be plasmids, cosmids or viruses such as phage or retroviruses. The coding sequence of the polypeptide is placed under the control of an appropriate promoter, control elements and transcription terminator so that the nucleic acid sequence encoding the polypeptide is transcribed into RNA in the host cell transformed or transfected by the expression vector construct. The coding sequence may or may not contain a signal peptide or leader sequence for secretion of the polypeptide out of the host cell. Expression and purification of the polypeptides of the invention can be easily performed using methods well known in the art (for example as described in Sambrook et al. "Molecular Cloning- A Laboratory Manual, second edition 1989").

The vectors containing the DNA coding for the ERG8 polypeptides of the invention can be introduced (i.e transformed or transfected) into *E. coli*, *S. cerevisiae*, *Pichia pastoris* or any other suitable host to facilitate their manipulation (i.e. for mutagenesis, cloning or expression). Performance of the invention is neither dependent on nor limited to any particular strain of host cell or vector; those suitable for use in the invention will be apparent to, and a matter of choice for, the person skilled in the art.

Host cells transformed or transfected with a vector containing an ERG8 nucleotide sequence may be cultured under conditions suitable for the expression and recovery of the encoded proteins from the cell culture. Such expressed proteins/polypeptides may be secreted into the culture medium or they may be contained intracellularly depending on the sequences used, i.e. whether or not suitable secretion signal sequences were present.

The full-length native isolated *C. albicans* ERG8 protein (PMK enzyme) of the present invention, or a functional variant thereof, is useful as a target in biochemical assays, particularly for use in identifying inhibitors of the enzyme. However, to provide sufficient enzyme for a biochemical assays (for example, for use in a high throughput screen for enzyme inhibitors) the enzyme has to be expressed at high levels and it has to be purified. Two major constraints impair ERG8 expression and purification: (i) ERG8 is not expressed at high levels from *C. albicans*, and (ii) expression and protein purification methodology is not well advanced for *C. albicans*.

We have now been able to overcome these problems by controlled over-expression of the *C. albicans* ERG8 in a strain of *Saccharomyces cerevisiae*. *S. cerevisiae* is a model system for expression and purification of recombinant proteins. Use of *S. cerevisiae* to express *C. albicans* ERG8 means that transformation, expression and purification

methodology used to produce and isolate the ERG8 protein can follow published procedures. As stated above, the invention is not limited to use of *S. cerevisiae* as the host for expression of *C. albicans* ERG8.

According to a further aspect of the invention there is provided a host cell adapted to express *C. albicans* ERG8 polypeptide or a variant thereof. The yeast *S. cerevisiae* is the preferred host cell of choice. According to a further aspect of the invention there is provided a novel expression system for expression of the *C. albicans* ERG8 gene, which system comprises an *S. cerevisiae* host strain having the *C. albicans* ERG8 gene in place of the native ERG8 gene from *S. cerevisiae*, whereby the *C. albicans* ERG8 gene is expressed. Preferred *S. cerevisiae* strains include JK9-3Daα and its haploid segregants.

The *C. albicans* ERG8 gene is preferably over-expressed relative to the expression derived from its own promoter. This is conveniently achieved by replacing the *C. albicans* ERG8 promoter by a stronger and preferably inducible promoter such as the *S. cerevisiae* GAL1 promoter, alpha factor or alcohol oxidase (for reviews see Ausubel et al. "Current Protocols in Molecular Biology", John Wiley & Sons, New York.).

The novel expression system is conveniently prepared by transformation of a heterozygous ERG8 deletion strain of a convenient *S. cerevisiae* host by a suitable plasmid comprising the *C. albicans* ERG8 gene using methods well known in the art (Ito et al. J. Bacteriol. 153:163-168,1983; Schiestl and Grietz, Current Genetics 16:339-346,1989).

The plasmid comprising the *C. albicans* ERG8 represents a further aspect of the invention. Particularly suitable plasmids for expression of *C. albicans* ERG8 in *S. cerevisiae* include pYES2(Invitrogen) and plasmids derived from pYES2 carrying a native *S. cerevisiae* promoter such as the glyceraldehyde-3-dehydrogenase promoter.

The heterozygous ERG8 deletion strain of a diploid *S. cerevisiae* host is conveniently achieved by disruption preferably using an antibiotic resistance cassette such as the kanamycin resistance cassette described by Wach et al (Yeast. 10:1793-1808, 1994).

As described earlier, the *C. albicans* ERG8 enzyme may be used in biochemical assays to identify agents which modulate the activity of the enzyme. The design and implementation of such assays will be evident to the biochemist of ordinary skill. The enzyme may be used to turn over a convenient substrate whilst incorporating/losing a labelled component to define a test system. Test compounds are introduced into the test system and measurements made to

determine their effect on enzyme activity. Such assays are useful to identify inhibitors of the enzyme which may then prove valuable as antifungal agents.

Thus, in a further aspect of the invention we provide the use of a *C. albicans* ERG-8 gene and/or *C. albicans* PMK enzyme in an assay to identify inhibitors of the enzyme. In particular, we provide their use in pharmaceutical or agrochemical research.

Thus, according to a further aspect of the invention there is provided a method of identifying compounds that modulate, preferably inhibit, the activity of phosphomevalonate kinase (PMK), comprising, contacting a test compound with a polypeptide of the invention and determining the effect that the test compound has on the activity of the polypeptide.

The PMK (ERG8) protein catalyses the conversion of phosphomevalonate + ATP to pyrophosphomevalonate + ADP. By way of non-limiting example, the activity of the ERG8 enzyme may be determined by (i) measuring the increase in ADP production, (ii) by following the loss of ATP, or (iii) by monitoring transfer of radioactive label (i.e H^3, C^{14}, P^{32}) into phosphomevalonate.

A suitable assay that measures ADP production involves coupling the ADP produced by the action of PMK on phosphomevalonate + ATP substrate with pyruvate kinase and phosphoenolpyruvate to form pyruvate and ATP. The pyruvate is then reduced to lactate with lactate dehydrogenase which converts NADH to NAD. The production of NAD (directly linked to ADP production indicative of PMK action) is conveniently measured by detecting the change in absorbance at 340nm (NADH oxidation product). In this assay, test compounds that inhibit PMK activity are identified by determining the ability of a compound to inhibit PMK activity as assessed by a reduction in ADP production as gauged by a reduction in the production of NAD from NADH using pyruvate kinase and lactate dehydrogenase as coupling enzymes as described above. The person skilled in the art would be able to develop other assays for measuring PMK activity without inventive input.

ATP can be conveniently assayed using commercially available kits (i.e Boehringer Mannheim) to monitor luminescence resulting from oxidation of luciferin to luciferase (Ford et al. J. Biolumin. Chemilumin. 11:149-167, 1996).

A suitable reaction that measures the production of radioactively labelled phosphomevalonate involves incubation PMK enzyme with cofactors, substrate ATP and phosphomevalonate, one of which carries a radioactive label. After reaction, pyrophosphomevalonate can be resolved from unreacted substrate by high voltage electrophoresis at pH3.5 on 3MM paper and the amount of radioactivity incorporated into

pyrophosphomevalonate can be measured by scintillation counting (Lee andO'Sullivan. J. Biol. Chem. 260:13909-13915, 1985).

Any convenient test compound or library of test compounds may be used in conjunction with the test assay. Particular test compounds include low molecular weight chemical compounds (preferably with a molecular weight less than 1500 daltons) suitable as pharmaceutical or veterinary agents for human or animal use, or compounds for non-administered use such as cleaning/sterilising agents or for agricultural use.

The ERG8 enzyme of the invention, and convenient fragments thereof may be used to raise antibodies. Such antibodies have a number of uses which will be evident to the molecular biologist or immunologist of ordinary skill. Such uses include, but are not limited to, monitoring enzyme expression, development of assays to measure enzyme activity, precipitation or purification of the enzyme and as a diagnostic tool to detect *C. albicans*. Enzyme linked immunosorbant assays (ELISAs) are well known in the art and would be particularly suitable for detecting the ERG8 polypeptide or fragments thereof. Antibodies raised against the polypeptides of the invention may be polyclonal, obtained for example by injecting the polypeptide(s) into a selected mammal (i.e. rabbit, mouse, goat or horse), and later collecting the immunised serum from the animal, and treating this according to procedures known in the art. Depending on the host species, various adjuvants may be used to enhance the immunological response against the injected polypeptide. Suitable adjuvants include, but are not limited to Freud's, aluminium hydroxide and SAF. Antibodies may also be monoclonal antibodies produced by hybridoma cells, phage display libraries or other methodology. Monoclonal antibodies may be inter alia, human, rat or mouse derived. For the production of human monoclonal antibodies, hybridoma cells may be prepared by fusing spleen cells from an immunised animal, e.g. a mouse, with a tumour cell. Appropriately secreting hybridoma cells may thereafter be selected (Koehler & Milstein. Nature. 256:495-497, 1975; Cole et al. "Monoclonal antibodies and Cancer Therapy, Alan R Liss Inc, New York N.Y. pp 77-96). Rodent antibodies may be humanised using recombinant DNA technology according to techniques known in the art. Alternatively, chimeric antibodies, single chain antibodies, Fab fragments may also be developed against the polypeptides of the invention (Huse et al. Science. 256:1275-1281, 1989), using skills known in the art.

The polynucleotides and antibodies of the invention may be used in gene-probe or protein-probe methodologies, with or without amplification (for example, via PCR or second antibody detection) to detect or diagnose the presence of *C. albicans*. This is particularly

valuable in diagnosing clinical infections. Accordingly, the invention provides diagnostic kits for the detection of *C. albicans* ERG8 or fragments thereof, and provides for the use of ERG8 protein, polypeptide fragments thereof and/or antibodies raised thereagainst as positive control. The reagents in the kit may be compartmentalised and the kit may also comprise instructions for use.

DNA diagnostics is based on DNA/RNA hybridisation technology, i.e. the specific *in vitro* binding of complementary single-stranded nucleic acid with the formation of double-stranded nucleic acid. The DNA/DNA or DNA/RNA double strands formed are termed hybrids. To detect the presence of *C. albicans* in a bodily fluid such as blood, total nucleic acid is isolated from the test fluid sample using standard techniques and the presence of *C. albicans* ERG8 nucleic acid in the sample is detected using for example detectably labelled probes comprising one or more of the polynucleotides of the invention. The probes can be short, chemically synthesised oligonucleotide probes of a length of approximately 10 - 50 nucleotides, or may be recombinantly expressed fragments of the ERG8 gene of approximately 0.3-1.5Kb in size. Single stranded oligonucleotide probes which are specific for *C. albicans* are preferred. The probe can be provided with a suitable detectable reporter molecule label such as a radioisotope (P^{32}, tritium, C^{14} or S^{35}), or a non-radioactive label such as digoxigenin or biotin, using techniques available to the person skilled in the art. Prior to the hybridisation reaction, all or any part of *C. albicans* ERG8 DNA containing the sequence to which the probe can hybridise, present in the test sample is amplified using for example PCR (polymerase chain reaction) or LCR (ligase chain reaction). For the specific hybridisation reaction, the test nucleic acid and if necessary the probe DNA is converted into single strands by denaturation (heat or alkali) and then very specifically hybridised with each other under stringent conditions. Under appropriate conditions the gene probe only hybridises to complementary sequences of the DNA or RNA to be detected. The hybridisation and detection assay can be carried out in a number of different formats known to the person skilled in the art including, solid-phase hybridisation of target DNA or probe coupled to a solid support such as nitrocellulose or magnetic beads. The hybridisation complex can then be determined quantitatively, following removal of unbound probe or test nucleic acid, by way of the reporter molecule label (e.g. fluorescent or radioactive) employed.

The test sensitivity of this single gene-probe diagnostic method can be increased by combination with DNA or RNA amplification techniques such as PCR or LCR. Using such amplification techniques, the DNA to be detected can be multiplied by up to 10^6.

There may only be 100-1000 organisms per ml of blood in association with Candida infections. Such small numbers of cells are easily detectable when combining the amplification and DNA-probe detection techniques offering the possibility of early detection of infection.

Thus, according to a further aspect of the invention there is provided a method of diagnosing the presence of the C. albicans ERG8 gene in a test sample, comprising: contacting a polynucleotide probe of at least 15 nucleotides in length, which probe is capable of specifically hybridising with the sequence depicted in SEQ ID No. 6, with the test sample under conditions which allow duplex formation between said polynucleotide probe and the nucleic acid in the test sample; and, detecting duplex formation. In a preferred embodiment the polynucleotide probe is detectably labelled. In another embodiment the polynucleotide probe is single stranded. In another embodiment the polynucleotide probe is completely complementary to the target sequence to be detected. According to a further aspect of the invention the polynucleotide probe is substituted for by a pair of oligonucleotide primers capable of specific PCR amplification of all or part of the ERG8 gene in the test sample, with subsequent identification of amplification product.

According to another aspect of the present invention there is provided a diagnostic kit for diagnosing or detecting the presence of C. albicans comprising, one or more diagnostic probe(s) and/or diagnostic primer(s) and/or antibodies capable of selectively hybridising or binding to the polynucleotide of SEQ ID No. 6 or the polypeptide of SEQ ID No. 7, or to variant sequences thereof as defined herein.

In a preferred embodiment, the diagnostic (detection) probes are provided on a microarray.

Such kits may further comprise appropriate buffer(s) and/or polymerase(s) such as thermostable polymerases, for example taq polymerase. They may also comprise companion/constant primers and/or control primers or probes. A companion/constant primer is one that is part of the pair of primers used to perform PCR. Such primer usually complements the template strand precisely.

In another embodiment the kit is an ELISA kit comprising one or more antibodies specific for the polypeptide depicted in SEQ ID No. 7, or a variant thereof as defined herein.

The following examples and figure describe and illustrate the invention. They are not intended to limit the scope of the invention in any way:

Figure 1 shows the nucleotide sequence of the *C. albicans* gene encoding phosphomevalonate kinase. Translation start (ATG) and stop (TAA) codons are highlighted.

Examples

1. Cloning and partial sequence determination of two separate clones from a *Candida albicans* genomic library.

Two separate cloned and sequenced nucleic acid sequences from a *C. albicans* library (SEQ ID NOs. 1 & 3) were found to have homology to that of *S. cerevisiae* ERG8 gene. The complement of specific regions in SEQ ID Nos. 1 and 3 were synthesised as oligonucleotides (SEQ ID Nos. 2 & 4) for use in the isolation of a clone containing the *C. albicans* ERG8 gene.

2. Cloning and sequence determination of *Candida albicans* ERG8.

Using the two oligonucleotide primers (SEQ ID Nos. 2 and 4), the *C. albicans* ERG8 gene was isolated as a plasmid clone from a library of *C. albicans* genomic DNA in the yeast shuttle vector YEp24 using PCR. The *C. albicans* library was maintained in *E. coli* and independent bacterial colonies were grown in single wells of each of 15 x 384-well microtitre plates. The properties of the library plasmids are such that this gridded array contains approximately 2.5x the amount of DNA in the *C. albicans* genome.

Small aliquots of cells from each of the wells were mixed to produce a pool of cells that were derived from all of the wells from a single plate. Similar pools were made for all of the rows and all of the columns from each of the plates. Samples of each of the pools of the cells for each complete plate were used in PCR reactions with SEQ ID Nos. 2 and 4 oligonucleotide primers to identify plate(s) in the array carrying *C. albicans* ERG8. Subsequent PCR reactions with pools of cells from rows of wells and columns of wells defined the specific well(s) carrying a clone of *C. albicans* ERG8.

The PCR reactions contained in a total volume of 0.05ml: 75mM Tris-HCl (pH 8.8 at 25°C), 20mM $(NH_4)_2SO_4$, 1.5mM $MgCl_2$, 0.01% Tween 20, 0.2mM of each of dATP, dCTP, dGTP and dTTP, 1.25 units Taq DNA polymerase, 100pmoles of each oligonucleotide primer and 0.005ml *E. coli* cell suspension. PCR reactions were incubated at 94°C for 1 min then for 30 cycles of the following: 94°C for 1 min, 55°C for 1 min, 72°C for 1 min. PCR products were analysed by electrophoresis through agarose and visualised under UV light after staining with ethidium bromide.

Putative clones harbouring the ERG8 gene were selected, the plasmid DNAs in these clones were purified and the complete sequence of the *C. albicans* ERG8 gene was determined on both strands using flanking sequence- or insert sequence-specific oligonucleotide primers. The full-length of the *C. albicans* ERG8 gene, including 'start' ATG and 'stop' TAA is shown in Figure 1. The protein translation of the gene is depicted in SEQ ID No. 7.

3. Generation of a heterozygous ERG8 deletion strain of *S. cerevisiae*

Since PMK is an essential enzyme, only one allele of a diploid cell can be deleted without loss of viability. One ERG8 gene diploid strain of *S. cerevisiae* (JK9-3daa; Kunz et al., Cell 73:585-596 (1993)) was disrupted using a kanamycin resistance cassette as described by Wach et al. (Yeast 10:1793-1808, 1994) using the protocol described therein with the oligonucleotides shown in SEQ ID Nos. 8 and 9. Sporulation of the heterozygous diploid (ERG8/erg8::KanMX) yields only two viable spores that are both sensitive to kanamycin, showing ERG8 to be essential, and the characteristic arrest phenotype for the two inviable spores.

4. Complementation of a *S. cerevisiae* ERG8 deletion with the cloned *C. albicans* ERG8

The heterozygous ERG8/erg8::KanMX strain was transformed with the plasmid carrying the full-length *C. albicans* ERG8 gene within a fragment of *C. albicans* genomic DNA such that expression of the gene will depend on functionality of the *C. albicans* promoter in the heterologous *S. cerevisiae* host. Surprisingly, the gene carried on the plasmid failed to complement the gene deletion as demonstrated by a failure to recover kanamycin-resistant haploid cells after sporulation. This was probably due to inappropriate expression of *C. albicans* ERG8 in *S. cerevisiae*.

To enable expression of *C. albicans* ERG8 in *S. cerevisiae* and to facilitate purification of ERG8 protein as a result of over-expression in a suitable host, the *C. albicans* promoter was replaced by the efficient, inducible *S. cerevisiae* GAL1 promoter. The *C. albicans* ERG8 coding sequence was amplified by PCR using the oligonucleotides shown in SEQ ID Nos. 10 and 11, which contain convenient restriction enzyme sites for cloning the product of PCR into an appropriate expression vector such as pYES2 (Invitrogen). The identity of the PCR-amplified gene cloned into pYES2 was confirmed by DNA sequencing. After transformation into the heterozygous *ERG8/erg8::KanMX* strain, the plasmid was able

to complement the erg8::KanMX allele in *S. cerevisiae* since kanamycin-resistant haploid spores were viable on medium containing galactose but not glucose. This *S. cerevisiae* strain is a useful source of biologically active *C. albicans* ERG8 protein for assays *in vitro*.

C. albicans ERG8 can also be conveniently over-expressed in bacteria such as *E. coli*. The *C. albicans* ERG8 coding sequence is amplified by PCR using oligonucleotides containing convenient restriction sites for cloning into expression vectors such as pT7#3.3. It is particularly convenient if the initiation codon for ERG8 is incorporated within one of the restriction sites. Oligonucleotides suitable for this are shown in SEQ ID Nos. 12 and 13. Oligonucleotides may also incorporate extra sequences to encode a small "tag" that aids the subsequent purification of the protein. Such tags include for example, the "His_6" tags which may be incorporated at the N- or C-terminus of ERG8 using the oligonucleotides shown in SEQ ID Nos. 14 and 15. Recombinantly expressed tagged ERG8 protein can be conveniently purified by affinity chromatography purification methodology using commercially available purification kits (i.e Qiagen) (Borsig et al., Biochem. Biophys. Res. Commun. 240:586-589, 1997).

5. Hybridisation test of nucleic acid variations of specific nucleic acid sequences

5.1 Hybridisation Test

A method for detecting variant nucleic acids containing sequences related to specific ERG8 sequences such as natural alleles, is described. These variant nucleic acids may be present in a variety of forms such as within plasmids or other like vehicles which may be fixed on to a hybridisation membrane, such as a nitrocellulose or nylon filter ready for detection using a labelled probe. Hybridisation assays can also be performed to identify variant sequences from within genomic or cDNA libraries. Hybridisation technology is well advanced. It will be apparent to the person skilled in the art that the protocol described below is only one example of a hybridisation protocol suitable to identify ERG8 variant sequences.

5.2 Hybridisation probe

Hybridisation probes may be generated from any fragment of DNA or RNA encoding the specific ERG8 nucleic sequence of interest. Such fragments can be for example, restriction fragments isolated following restriction enzyme digestion of nucleic acid containing the ERG8 nucleotide sequence or synthetic oligonucleotides specific for a region of the ERG8 gene or a complementary sequence thereto.

A hybridisation probe can be generated from a synthetic oligonucleotide or a dephosphorylated restriction fragment sequence by addition of a radioactive 5' phospate group from [γ-^{32}P]ATP by the action of T4 polynucleotide kinase. 20 pmoles of the oligonucleotide are added to a 20µl reaction containing 100mM Tris, pH7.5 , 10mM $MgCl_2$, 0.1mM spermidine, 20mM dithiothreitol (DTT), 7.55µM ATP, 55µCi [γ-^{32}P]ATP and 2.5u T4 polynucleotide kinase (Pharmacia Biotechnology Ltd, Uppsala, Sweden). The reaction is incubated for 30 minutes at 37°C and then for 10 minutes at 70°C prior to use in hybridisation. Methods for the generation of hybridisation probes from oligonucleotides or from DNA and RNA fragments (Chapters 11 and 10 respectively in Sambrook et al. *ibid*). A number of proprietary kits are also available for these procedures.

5.3 Hybridisation conditions

Filters containing the nucleic acid are pre-hybridised in 100ml of a solution containing 6x SSC, 0.1%SDS and 0.25% dried skimmed milk (Marvel™) at 65°C for a minimum of 1 hour in a suitable enclosed vessel. A proprietary hybridisation apparatus such as model HB-1 (Techne Ltd) provides reproducible conditions for the experiment.

The pre-hybridisation solution is then replaced by 10ml of a probe solution containing 6xSSC, 0.1% SDS, 0.25% dried skimmed milk (e.g. Marvel™) and the oligonucleotide probe generated above. The filters are incubated in this solution for 5 minutes at 65°C before allowing the temperature to fall gradually to below 30°C. The probe solution is then discarded and the filters washed in 100ml 6xSSC, 0.1% SDS at room temperature for 5 minutes. Further washes are then made in fresh batches of the same solution at 30°C and then in 10°C increments up to 60°C for 5 minutes per wash.

After washing, the filters are dried and used to expose an X-ray film such as Hyperfilm™ MP (Amersham International) at -70°C in a light-tight film cassette using a fast tungstate intensifying screen to enhance the photographic image. The film is exposed for a suitable period (normally overnight) before developing to reveal the photographic image of the radio-active areas on the filters. Related nucleic acid sequences are identified by the presence of a photographic image compared to totally unrelated sequences which should not produce an image. Generally, related sequences will appear positive at the highest wash

temperature (60°C). However, related sequences may only show positive at the lower wash temperatures (50, 40 or 30°C).

These results will also depend upon the nature of the probe used. Longer nucleic acid fragment probes will need to be hybridised for longer periods at high temperature but may remain bound to related sequences at higher wash temperatures and/or at lower salt concentrations. Shorter, mixed or degenerate oligonucleotide probes may require less stringent washing conditions such as lower temperatures and/or higher Na^+ concentrations. A discussion of the considerations for hybridisation protocols is provided in Sambrook et al. (Chapter 11).

To prepare 20 x SSC,175.3 g of NaCl and 88.2 g of sodium citrate is dissolved in approximately 800ml of water, the pH is adjusted to 7.0 using 10 N solution of NaOH and the volume is adjusted to 1 litre with water, before autoclaving.

Claims:

1. A purified polypeptide comprising the amino acid sequence depicted in SEQ ID No. 7 or a sequence possessing at least 80% similarity thereto.
2. An isolated polypeptide of at least 15 contiguous amino acids of the polypeptide of claim 1.
3. An antibody specific for the polypeptide of claim 1 or 2.
4. An antibody as claimed in claim 3 which is a monoclonal antibody.
5. A purified polynucleotide comprising a nucleic acid sequence encoding the polypeptide depicted in SEQ ID No. 7 or a sequence possessing at least 80% identity thereto.
6. A polynucleotide of at least 15 nucleotides in length, which polynucleotide is capable of specifically hybridising to a nucleic acid sequence selected from the group consisting of SEQ ID Nos. 1, 3, 5 or 6, or a sequence complementary to any of said sequences.
7. An expression vector comprising the polynucleotide of claim 5.
8. A host cell which contains an expression vector according to claim 7.
9. A method for producing the polypeptide of claim 1, comprising:

(a) culturing a host cell according to claim 8 under conditions suitable for the expression of said polypeptide, and

(b) recovering said polypeptide from the host cell or cell culture.

10. Use of the polypeptide of claim 1 in an assay to identify compounds that inhibit phosphomevalonate kinase (PMK) activity.
11. A method of identifying compounds that modulate the activity of PMK, comprising:

(a) contacting a test compound with a polypeptide according to claim 1, and

(b) determining the effect that the test compound has on the activity of the polypeptide.

12. A compound identified by the method of claim 11.
13. A method for detecting or diagnosing the presence of *Candida albicans* in a test sample, comprising contacting the sample with an agent capable of detecting a polypeptide possessing the amino acid sequence depicted in SEQ ID No. 7 or a sequence possessing at least 80% similarity thereto, or a nucleic acid sequence encoding the polypeptide depicted in SEQ ID No. 7 or a sequence possessing at least 80% identity thereto.
14. A method as claimed in claim 13 wherein the presence of the nucleic acid is detected using an oligonucleotide primer or probe capable of selectively hybridising to the said polynucleotide.
15. A diagnostic kit for detecting the presence of C. albicans comprising: one or more diagnostic probe(s) and/or diagnostic primer(s) and/or antibodies capable of selectively

hybridising or binding to the polynucleotide of claim 6 or the polypeptide of claim 1, and instructions for use.

GTGGAAAAAAAAGACAGAACAGTAGATTCCAACTTCAGAATATTCATTCAGATCTGAACATTT
CTTTTTCTCCGATCATCAATTGGCAATGTCAAAAGCATTTAGTGCACCTGGAAAAGCATTTCT
TGCTGGTGGATATTTGGTTCTTGAGCCAATTTATGATGCTTATGTGACAGCATTGTCATCACG
AATGCATGCAGTTATAACACCAAAAGGAACCAGTTTGAAAGAATCTAGAATCAAAATTTCTTC
ACCCCAATTTGCAAACGGAGAATGGGAATATCACATATCATCAAATACAGAGAAGCCCAGAGA
AGTTCAGTCACGCATAAATCCATTTTTAGAGGCAACTATATTCATCGTTTTAGCTTATATTCA
ACCGACCGAAGCATTTGATCTTGAAATCATCATTTACTCAGACCCTGGATATCATTCACAAGA
AGATACTGAAACCAAGACATCCTCGAATGGAGAAAAAACATTTCTTTACCATTCTCGTGCCAT
TACCGAAGTGGAAAAGACCGGATTAGGTTCATCGGCAGGATTAGTGTCAGTTGTTGCCACAAG
TTTATTATCCCATTTTATCCCCAATGTTATCAGTACGAATAAAGATATTTTGCACAACGTTGC
ACAGATTGCACATTGTTATGCCCAAAAAAAGATAGGATCTGGGTTTGATGTTGCAACTGCAAT
TTATGGTCTGATTGTATATAGAAGATTTCAGCCAGCTTTGATAAATGACGTGTTTCAGGTTCT
AGAAAGTGATCCTGAGAAGTTCCCCACAGAGTTGAAAAAATTGATTGAAAGTAACTGGGAATT
CAAACATGAAAGATGTACATTACCATACGGAATCAAGTTATTAATGGGTGACGTCAAGGGTGG
CTCAGAAACACCCAAATTGGTATCACGAGTACTCCAATGGAAAAAGGAAAAGCCAGAAGAAAG
CTCTGTTGTGTATGACCAGCTTAATAGTGCCAATTTACAGTTTATGAAGGAATTGAGGGAAAT
GCGTGAAAAATACGACTCAGACCCAGAGACTTATATTAAAGAGTTAGATCATTCTGTTGAGCC
TTTGACTGTTGCGATTAAGAACATCAGAAAAGGGTTACAAGCATTAACACAAAAATCAGAGGT
TCCAATTGAACCTGATGTCCAAACCCAGTTGTTGGACCGTTGTCAAGAGATTCCTGGTTGTGT
TGGTGGTGTGGTTCCAGGTGCTGGTGGATACGATGCAATAGCTGTATTAGTGTTGGAAAATCA
AGTGGGAAATTTTAAGCAGAAAACTCTTGAAAATCCAGATTATTTTCATAATGTTTACTGGGT
TGATTTGGAAGAGCAAACAGAAGGTGTACTTGAAGAAAAACCAGAAGACTATATAGGTTTATA
AAATATCACTGGGATATGTCTACAAGGTGTTTTCGATTAGAGTTTTTGATCCCCATTTTAACA
TATTTTACTTCAATCTTACACTTTATCCTTTTAAGTAGGTATGTGTAGGGAAAGAGCCTGATC
TTCATAAACCGTTGCAAACTAATTGATTATATTTTCTATTGTAAATTTCATATGCAGGAAATA
GCTTATTCGACAAATTATTTATTTTCGTCTCGTTCTGGTCCAAGTACCCCAGAGACGAAATAA
CTGACAACACGCAGGGCTGGGTTGGCATTTTCGTCACACGATTATTATTAATGGTAACAAAAA
AAGGGGRKATGCCCGTGGTCGATACACAAATATTTATGATATACTTTCCATATTTTTTTTTT

FIGURE 1

SEQUENCE LISTING

<110> ASTRAZENECA AB

<120> PROTEIN

<130> LDSG/PHM70579/WO

<140>

<141>

<150> GB 9919766.7

<151> 1999-08-21

<160> 11

<170> PatentIn Ver. 2.1

<210> 1

<211> 547

<212> DNA

<213> Candida albicans

<400> 1

ccaatggaaa aaggaaaagc cagaagaaag ctctgttgtg tatgaccagc ttaatagtgc 60
caatttacag tttatgaagg aattgaggga aatgcgtgaa aaatacgact cagacccaga 120
gacttatatt aaagagttag atcattctgt tgagcctttg actgttgcga ttaagaacat 180
cagaaaaggg ttacaagcat taacacaaaa atcagaggtt ccaattgaac ctgatgtcca 240
aacccagttg ttggaccgtt gtcaagagat tcctggttgt gttggcggtg tggttccagg 300
tgctggtgga tacgatgcaa tagctgtatt agtgttggaa aatcaagtgg gaaattttaa 360
gcagaaaact cttgaaaatc cagattattt tcataatgtt tactgggttg atttggaaga 420
gcaaacagaa ggtgtacttg aagaaaaacc agaagactat ataggtttat aaaatatcac 480
taggatatgt ctacaaggtg atttcgatta gattttctgc tacccgtttt aacatatttt 540
acttcaa 547

<210> 2

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Single-stranded oligonucleotide

<400> 2

gctggtggat acgatgcaat a 21

<210> 3
<211> 577
<212> DNA
<213> Candida albicans

<400> 3
atgtacatct ttcatgtttg aattcccagt tacttgcaat caatttttc aactctgtgg 60
ggaacttctc aggatcactt tctagaacct gaaacacgtc atttatcaaa gctggctgaa 120
atcttctata tacaatcaga ccataaattg cagttgcaac atcaaaccca gatcctatct 180
ttttttgggc ataacaatgt gcaatctgtg caacgttgtg caaaatatct ttattcgtac 240
tgataacatt ggggataaaa tgggataata aacttgtggc aacaactgac actaatcctg 300
ccgatgaacc taatccggtc ttttccactt cggtaatggc acgagaatgg taaagaaaag 360
ttttttctcc attcgaggat gtcttggttt cagtatcttc ttgtgaatga tatccagggt 420
ccgagtaaat aatgatttca agatcaaatg cttcggtcgg ttgaatataa gctaaaaacc 480
gatggatata gttgcctcta aaaatgggat ttatgcgtga ctgnacttct ttgggttttc 540
ngtaattgat gatatgtgat antcccattc cccggtt 577

<210> 4
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Single-stranded oligonucleotide

<400> 4
ggggataaaa tgggataata aactt 25

<210> 5
<211> 1763
<212> DNA
<213> Candida albicans

<400> 5
gtggaaaaaa aagacagaac agtagattcc aacttcagaa tattcattca gatctgaaca 60
tttctttttc tccgatcatc aattggcaat gtcaaaagca tttagtgcac ctggaaaagc 120
atttcttgct ggtggatatt tggttcttga gccaatttat gatgcttatg tgacagcatt 180
gtcatcacga atgcatgcag ttataacacc aaaaggaacc agtttgaaag aatctagaat 240
caaaatttct tcaccccaat ttgcaaacgg agaatgggaa tatcacatat catcaaatac 300
agagaagccc agagaagttc agtcacgcat aaatccattt ttagaggcaa ctatattcat 360
cgttttagct tatattcaac cgaccgaagc atttgatctt gaaatcatca tttactcaga 420
ccctggatat cattcacaag aagatactga aaccaagaca tcctcgaatg gagaaaaaac 480
atttctttac cattctcgtg ccattaccga agtggaaaag accggattag gttcatcggc 540
aggattagtg tcagttgttg ccacaagttt attatcccat tttatcccca atgttatcag 600
tacgaataaa gatattttgc acaacgttgc acagattgca cattgttatg cccaaaaaaa 660
gataggatct gggtttgatg ttgcaactgc aatttatggt ccgaccggac aaagaaggct 720

tcagccagct ttgataaatg acgtgtttca ggttctagaa agtgatcctg agaagttccc 780
cacagagttg aaaaaattga ttgaaagtaa ctgggaattc aaacatgaaa gatgtacatt 840
accatacgga atcaagttat taatgggtga cgtcaagggt ggctcagaaa cacccaaatt 900
ggtatcacga gtactccaat ggaaaaagga aaagccagaa gaaagctctg ttgtgtatga 960
ccagcttaat agtgccaatt tacagtttat gaaggaattg agggaaatgc gtgaaaaata 1020
cgactcagac ccagagactt atattaaaga gttagatcat tctgttgagc ctttgactgt 1080
tgcgattaag aacatcagaa aagggttaca agcattaaca caaaaatcag aggttccaat 1140
tgaacctgat gtccaaaccc agttgttgga ccgttgtcaa gagattcctg gttgtgttgg 1200
tggtgtggtt ccaggtgctg gtggatacga tgcaatagct gtattagtgt tggaaaatca 1260
agtgggaaat tttaagcaga aaactcttga aaatccagat tattttcata atgtttactg 1320
ggttgatttg gaagagcaaa cagaaggtgt acttgaagaa aaaccagaag actatatagg 1380
tttataaaat atcactggga tatgtctaca aggtgttttc gattagagtt tttgatcccc 1440
attttaacat attttacttc aatcttacac tttatccttt taagtaggta tgtgtaggga 1500
aagagcctga tcttcataaa ccgttgcaaa ctaattgatt atattttcta ttgtaaattt 1560
catatgcagg aaatagctta ttcgacaaat tatttatttt cgtctcgttc tggtccaagt 1620
accccagaga cgaaataact gacaacacgc agggctgggt tggcattttc gtcacacgat 1680
tattattaat ggtaacaaaa aaaggggrka tgcccgtggt cgatacacaa atatttatga 1740
tatactttcc atatttttt ttt 1763

<210> 6
<211> 1299
<212> DNA
<213> Candida albicans

<400> 6
atgtcaaaag catttagtgc acctggaaaa gcatttcttg ctggtggata tttggttctt 60
gagccaattt atgatgctta tgtgacagca ttgtcatcac gaatgcatgc agttataaca 120
ccaaaaggaa ccagtttgaa agaatctaga atcaaaattt cttcacccca atttgcaaac 180
ggagaatggg aatatcacat atcatcaaat acagagaagc ccagagaagt tcagtcacgc 240
ataaatccat ttttagaggc aactatattc atcgttttag cttatattca accgaccgaa 300
gcatttgatc ttgaaatcat catttactca gaccctggat atcattcaca agaagatact 360
gaaaccaaga catcctcgaa tggagaaaaa acatttcttt accattctcg tgccattacc 420
gaagtggaaa agaccggatt aggttcatcg gcaggattag tgtcagttgt tgccacaagt 480
ttattatccc attttatccc caatgttatc agtacgaata aagatatttt gcacaacgtt 540
gcacagattg cacattgtta tgcccaaaaa aagataggat ctgggtttga tgttgcaact 600
gcaatttatg gtctgattgt atatagaaga tttcagccag ctttgataaa tgacgtgttt 660
caggttctag aaagtgatcc tgagaagttc cccacagagt tgaaaaaatt gattgaaagt 720
aactgggaat tcaaacatga aagatgtaca ttaccatacg gaatcaagtt attaatgggt 780
gacgtcaagg gtggctcaga aacacccaaa ttggtatcac gagtactcca atggaaaaag 840
gaaaagccag aagaaagctc tgttgtgtat gaccagctta atagtgccaa tttacagttt 900
atgaaggaat tgagggaaat gcgtgaaaaa tacgactcag acccagagac ttatattaaa 960
gagttagatc attctgttga gcctttgact gttgcgatta agaacatcag aaaagggtta 1020
caagcattaa cacaaaaatc agaggttcca attgaacctg atgtccaaac ccagttgttg 1080
gaccgttgtc aagagattcc tggttgtgtt ggtggtgtgg ttccaggtgc tggtggatac 1140
gatgcaatag ctgtattagt gttggaaaat caagtgggaa attttaagca gaaaactctt 1200
gaaaatccag attattttca taatgtttac tgggttgatt tggaagagca aacagaaggt 1260
gtacttgaag aaaaaccaga agactatata ggttttataa 1299

<210> 7
<211> 432
<212> PRT
<213> Candida albicans

<400> 7

Met Ser Lys Ala Phe Ser Ala Pro Gly Lys Ala Phe Leu Ala Gly Gly
1 5 10 15

Tyr Leu Val Leu Glu Pro Ile Tyr Asp Ala Tyr Val Thr Ala Leu Ser
20 25 30

Ser Arg Met His Ala Val Ile Thr Pro Lys Gly Thr Ser Leu Lys Glu
35 40 45

Ser Arg Ile Lys Ile Ser Ser Pro Gln Phe Ala Asn Gly Glu Trp Glu
50 55 60

Tyr His Ile Ser Ser Asn Thr Glu Lys Pro Arg Glu Val Gln Ser Arg
65 70 75 80

Ile Asn Pro Phe Leu Glu Ala Thr Ile Phe Ile Val Leu Ala Tyr Ile
85 90 95

Gln Pro Thr Glu Ala Phe Asp Leu Glu Ile Ile Ile Tyr Ser Asp Pro
100 105 110

Gly Tyr His Ser Gln Glu Asp Thr Glu Thr Lys Thr Ser Ser Asn Gly
115 120 125

Glu Lys Thr Phe Leu Tyr His Ser Arg Ala Ile Thr Glu Val Glu Lys
130 135 140

Thr Gly Leu Gly Ser Ser Ala Gly Leu Val Ser Val Val Ala Thr Ser
145 150 155 160

Leu Leu Ser His Phe Ile Pro Asn Val Ile Ser Thr Asn Lys Asp Ile
165 170 175

Leu His Asn Val Ala Gln Ile Ala His Cys Tyr Ala Gln Lys Lys Ile
180 185 190

Gly Ser Gly Phe Asp Val Ala Thr Ala Ile Tyr Gly Leu Ile Val Tyr
195 200 205

Arg Arg Phe Gln Pro Ala Leu Ile Asn Asp Val Phe Gln Val Leu Glu

210 215 220

Ser Asp Pro Glu Lys Phe Pro Thr Glu Leu Lys Lys Leu Ile Glu Ser
225 230 235 240

Asn Trp Glu Glu Lys His Glu Arg Cys Thr Leu Pro Tyr Gly Ile Lys
245 250 255

Leu Leu Met Gly Asp Val Lys Gly Gly Ser Glu Thr Pro Lys Leu Val
260 265 270

Ser Arg Val Leu Gln Trp Lys Lys Glu Lys Pro Glu Glu Ser Ser Val
275 280 285

Val Tyr Asp Gln Leu Asn Ser Ala Asn Leu Gln Phe Met Lys Glu Leu
290 295 300

Arg Glu Met Arg Glu Lys Tyr Asp Ser Asp Pro Glu Thr Tyr Ile Lys
305 310 315 320

Glu Leu Asp His Ser Val Glu Pro Leu Thr Val Ala Ile Lys Asn Ile
325 330 335

Arg Lys Gly Leu Gln Ala Leu Thr Gln Lys Ser Glu Val Pro Ile Glu
340 345 350

Pro Asp Val Gln Thr Gln Leu Leu Asp Arg Cys Gln Glu Ile Pro Gly
355 360 365

Cys Val Gly Gly Val Val Pro Gly Ala Gly Gly Tyr Asp Ala Ile Ala
370 375 380

Val Leu Val Leu Glu Asn Gln Val Gly Asn Phe Lys Gln Lys Thr Leu
385 390 395 400

Glu Asn Pro Asp Tyr Phe His Asn Val Tyr Trp Val Asp Leu Glu Glu
405 410 415

Gln Thr Glu Gly Val Leu Glu Glu Lys Pro Glu Asp Tyr Ile Gly Leu
420 425 430

<210> 8
<211> 70
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Single-stranded oligonucleotide

<400> 8
aaatgtcaga gttgagagcc ttcagtgccc cagggaaagc gttactagct gcagctgaag 60
cttcgtacgc 70

<210> 9
<211> 73
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Single-stranded oligonucleotide

<400> 9
agttatttat caagataagt ttccggatct ttttctttcc taacacccca ggcataggcc 60
actagtggat ctg 73

<210> 10
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Single-stranded oligonucleotide

<400> 10
cccaagcttg gcaatgtcaa aagcatttag tgc 33

<210> 11
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Single-stranded oligonucleotide

<400> 11
ccgctcgaga ttttataaac ctatatagtc ttctgg 36